ADOLFO FIGUEROA

NUESTRO MUNDO SOCIAL

INTRODUCCIÓN A LA CIENCIA ECONÓMICA

FONDO EDITORIAL
PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

Adolfo Figueroa

Nuestro mundo social

Introducción a la ciencia económica

FONDO EDITORIAL
PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

Nuestro mundo social
Introducción a la ciencia económica
Primera edición, marzo de 2008

de la Pontificia Universidad Católica del Perú, 2008
Av. Universitaria 1801, Lima 32, Perú
Teléfono: (51 1) 626-2000
feditor@pucp.edu.pe
www.pucp.edu.pe/publicaciones

Diseño de cubierta: Roberto C. Torres Mautino
Diagramación de interiores: Aída Nagata

ISBN: 978-9972-42-844-9
Hecho el Depósito Legal en la Biblioteca Nacional
del Perú N° 2008-03893
Impreso en el Perú – Printed in Peru

Índice

Agradecimientos

En la preparación de este libro he recibido el apoyo de varias personas. Recibí de mis colegas del Departamento de Economía de la Pontificia Universidad Católica del Perú su cooperación de distintas formas y en distintas etapas. Los profesores Pedro Francke y Efraín Gonzales de Olarte me hicieron llegar comentarios valiosos a los primeras versiones del manuscrito; Giannina Vaccaro, como asistente de investigación, colaboró por varios meses en la preparación de bibliografía, cuadros y gráficos, así como en el diseño de un texto que fuese accesible al gran público; Mirtha Cornejo, secretaria del departamento, tuvo la paciencia de producir con gran destreza varias versiones del manuscrito.

Francisco Zela, profesor de física de esta misma universidad, tuvo la generosidad de darme sus comentarios al apéndice, en donde presento un análisis epistemológico sobre las diferencias en la naturaleza del conocimiento entre la economía y las ciencias naturales.

Ricardo Chuquín tuvo a bien entregarme comentarios y sugerencias a una versión inicial del manuscrito desde su experiencia de trabajo con organizaciones populares.

En el Fondo Editorial, su directora Patricia Arévalo me animó y embarcó en el proyecto de escribir este libro; Jenny Varillas, como editora, me ayudó con mucho profesionalismo en la difícil tarea de transformar un manuscrito en un libro.

Quiero expresar mi más profundo agradecimiento a todas estas personas.

Deseo igualmente agradecer a la Pontificia Universidad Católica del Perú y a su Departamento de Economía por haber creado un ambiente intelectual idóneo para el desarrollo de la investigación científica; de este ambiente me beneficié durante todos los años que estuve como profesor y pude desarrollar mis investigaciones teóricas y empíricas en la ciencia económica. Los aspectos fundamentales de mis hallazgos se recogen en este libro.

Presentación

Amigo lector, ¿no le gustaría entender cómo funciona el mundo social en que vivimos? Varios científicos de las ciencias naturales han escrito libros para el gran público, los que nos han permitido comprender mejor el mundo físico y el mundo biológico en que vivimos. Además, han elevado nuestro nivel de conciencia sobre varios problemas que atañen a la humanidad, como el del deterioro del medio ambiente.

Falta, sin embargo, una obra de alcance general que nos explique el mundo social en que vivimos, que nos responda la pregunta ¿cómo funciona nuestra sociedad? Este tipo de obra elevaría nuestro nivel de conocimiento sobre la realidad social y también elevaría nuestro nivel de conciencia sobre la naturaleza de los problemas sociales que confrontamos. Pero, al igual que en el caso de las ciencias naturales, tendría que basarse en el conocimiento científico.

Este libro, amigo lector, busca llenar ese vacío y cumplir con este requisito. Intenta explicar de manera sencilla cómo funciona el mundo social en que vivimos desde el punto de vista de una ciencia social particular: la ciencia económica. Busca mostrar cómo la ciencia económica puede ser utilizada para explicar la realidad social de manera sencilla, centrándose en sus aspectos fundamentales, pero manteniendo el rigor científico. Se puede, así, ver la ciencia económica en acción. Este libro es, en ese sentido particular, una introducción a la ciencia económica, ya que se dirige a dos públicos: al gran público lector y a los estudiantes universitarios.

Amigo lector, no deje que la ciencia lo intimide. Al contrario, trate de acercarse a ella. «La ciencia es una cosa maravillosa», como se dice usualmente. Para leer esta introducción a la ciencia económica el requisito fundamental es que usted tenga el interés de entender el mundo social en que vive. El dicho popular «A buen hambre no hay pan duro» se aplica muy bien en este caso. Para desarrollar su apetito tal vez sea importante señalar que la comprensión científica de la realidad social le será de mucha utilidad para interactuar con esta realidad; le ayudará a ser un ciudadano de primera categoría.

La lectura de este libro solo requiere conocimientos básicos sobre el manejo numérico, incluyendo cierta familiaridad con gráficos y cuadros. El manejo de fórmulas y ecuaciones complicadas, o matemáticas avanzadas, no constituyen requisito para comprender el contenido del texto. El físico inglés Stephen Hawking nos advierte que por cada ecuación que aparece en un libro se pierde diez por ciento de los potenciales lectores; en efecto, en su libro aparece una sola ecuación, la famosa ecuación de la energía de Einstein. Aquí vamos a tratar de respetar este principio. Las ecuaciones que se presentan son pocas y muy sencillas. Pero va una sugerencia, amigo lector: en cualquier caso, no se deje atemorizar por las matemáticas. El requisito más importante para leer este libro es, como mencionamos anteriormente, su interés por entender nuestra sociedad.

Para los estudiantes universitarios, dado que los fundamentos de la ciencia económica se presentan aquí, y se presentan en acción, el libro puede servir como texto para los cursos introductorios. A la vez que aprenden esos fundamentos, también podrán aprender y comprender el funcionamiento del mundo social en que vivimos. No hay que olvidar lo obvio: el objeto principal de una ciencia es ayudarnos a entender la realidad. Ojalá el libro pudiera despertar en los jóvenes aspiraciones para llegar a ser científicos sociales.

Sobre la naturaleza del libro

Sabemos que los humanos necesitamos bienes para satisfacer nuestras necesidades. También sabemos que algunos bienes son libres, pero que otros deben ser producidos. ¿Cómo se organizan las sociedades humanas para producir los bienes y para distribuirlos entre los distintos grupos sociales? Esta es la pregunta fundamental que la ciencia económica busca responder.

La ciencia económica tiene por objeto estudiar el proceso de producción y distribución de bienes, llamado el *proceso económico*. Se interesa por explicar cómo y por qué, en dicho proceso, los miembros de la sociedad entran en relaciones, interactúan entre ellos, y por qué lo hacen en diferentes grados y de diferentes maneras, jugando diferentes roles. Es por eso que la economía es una ciencia social.

El campo de estudio de la ciencia económica que está más desarrollado se refiere al mundo social en que vivimos: la sociedad capitalista. Esta realidad contemporánea será nuestro objeto de estudio. Para tal efecto, se presenta, de un lado, los hechos básicos de la sociedad capitalista de nuestro tiempo que la ciencia económica busca explicar; y, de otro, el método que sigue para llevar a cabo esa tarea.

En la ciencia económica se pueden distinguir dos tipos de preguntas: una se refiere a cómo *es* la realidad y la otra a cómo *debe ser* la realidad. Existen por lo tanto dos ramas científicas para responderlas: la ciencia económica positiva, que se dirige a responder la primera pregunta, y la ciencia económica normativa, que responde a la segunda. Este libro busca responder ambas preguntas y, por consiguiente, desarrollar ambas ramas científicas.

La economía como ciencia positiva necesita del uso de teorías para responder la pregunta *¿cómo es el mundo social en que vivimos?* Sin teoría no se puede explicar la realidad, ya que esta no es sino un artificio lógico que consiste en construir un mundo abstracto —que es mucho más simple— que se parezca al mundo real —que es mucho más complejo—. Ese parecido necesita ser corroborado estadísticamente, antes

de que la teoría sea aceptada como una *buena teoría*. Teoría y realidad irán de la mano en este libro. Otra sugerencia, amigo lector: tampoco se deje intimidar por el uso de teorías.

También trataremos la pregunta normativa que siempre nos hacemos: *¿cómo debería ser el mundo social en que vivimos?* Una cosa es que le digan a usted, amigo lector, cómo es el mundo, pero otra muy distinta es que le digan cómo debería ser el mundo. Por ejemplo, a algunos no les gusta el control de precios que hacen los gobiernos, pero a otros sí. A primera vista, no parece existir error posible entre estas posiciones de la gente, pues parece ser solamente un asunto de diferencias en sus preferencias.

Sin embargo, una proposición normativa tiene dos lados. El primero es, en efecto, asunto de preferencias o creencias. La gente gusta de algo porque simplemente le gusta. Esto es semejante a enunciar que a algunos les gusta el helado de vainilla y a otros el de chocolate. Y no se puede decir que alguien está equivocado. El segundo lado es lógico y aparece cuando es necesario justificar lo que a uno le gusta. Y esa justificación tiene que tener una lógica, una teoría normativa. Así surge la ciencia normativa. Los deseos, creencias y preferencias de la gente ya no serán suficientes para proponer cómo debe ser el mundo. Aquí sí aparece la posibilidad de error en una proposición normativa, pues la justificación, que necesariamente será una teoría positiva, puede ser empíricamente falsa. Una proposición normativa no puede ser un conjunto de buenos deseos sobre cómo debe ser la realidad, pues dejaría de tener utilidad.

En ese sentido, la ciencia normativa se justifica en la ciencia económica —y en las ciencias sociales en general— porque se considera que las relaciones entre personas en el mundo social son cambiables. En las ciencias naturales, por contraste, la pregunta normativa no tiene mucha importancia, pues las relaciones entre objetos en el mundo físico no son cambiables por nosotros, o lo son en menor grado. Por ejemplo, no podemos correr y levantar vuelo, como quisiéramos, pues hay relaciones entre los cuerpos en el mundo físico que nos lo impiden;

las cuatro estaciones del año existen independientemente de nuestros deseos, creencias y preferencias. Las relaciones que las ciencias naturales como ciencia positiva nos descubren son independientes de nuestros deseos. De manera similar, la economía como ciencia positiva nos mostrará relaciones en el mundo social que también son independientes de nuestros deseos, relaciones que pueden o no ser de nuestro agrado.

La distinción entre ciencia positiva y ciencia normativa es fundamental en la ciencia económica. Como ya hemos señalado, nuestros deseos sobre cómo debe ser el mundo requieren de una justificación lógica si nos van a servir para algo. Esta lógica está desarrollada en la ciencia económica normativa, este libro utilizará este tipo de ciencia para el análisis de las políticas públicas, que es el espacio donde los actores sociales discuten cómo debe ser el mundo social en que vivimos, utilizando para ello los conocimientos que nos entrega la ciencia positiva. Como ve, las dos ramas de la ciencia económica, la positiva y la normativa, irán aquí de la mano.

En suma, este es un libro sobre ciencia, en particular, sobre la ciencia económica que busca explicar, bajo las reglas de la ciencia positiva, el funcionamiento del mundo social, nos guste o nos disguste tal explicación. Incluso el tema de las políticas públicas para modificar la realidad social se presentará bajo las reglas de la ciencia normativa. Por lo tanto, los debates que puedan originar las conclusiones de esta obra tendrán que hacerse, necesariamente, siguiendo las reglas del conocimiento científico. Para tal efecto, dichas reglas se presentarán aquí de manera explícita y accesible, ya que el libro busca mostrara la ciencia económica en acción.

Sobre el contenido del libro

El contenido del libro se puede resumir en las preguntas específicas que se busca responder en cada capítulo. Estas son:

- ¿Cuál es el objeto de estudio de la ciencia económica y cómo lo lleva a cabo? Se necesita definir las cuestiones que estudia la economía y las que no estudia. Se necesita identificar el método científico que utiliza la economía para llegar a establecer una buena teoría que explique la realidad. El primer capítulo responde estas preguntas. Allí se establece un método científico particular que se utilizará a lo largo del texto. Este método establece un conjunto de reglas lógicas para llegar al conocimiento científico. Todas las conclusiones del libro resultan de la aplicación de estas reglas, las cuales conducen a un algoritmo lógico: no se puede dar un paso adelante si el paso anterior no está bien definido. Vuelva a este capítulo, amigo lector, cada vez que sienta que los pasos que se siguen en el texto para llegar a las conclusiones no le parezcan claros.

- Toda ciencia construye el conjunto de hechos que se repiten de manera sistemática y que luego constituye su agenda de investigación. ¿Cuáles son los hechos básicos del sistema capitalista que tienen que ver con la producción y la distribución de bienes? El segundo capítulo presenta las siete regularidades empíricas más importantes del funcionamiento del capitalismo contemporáneo. Para establecer dichas regularidades se divide el mundo capitalista en dos grupos, según su nivel de riqueza: el primer mundo y el tercer mundo. El objetivo central del libro es responder la pregunta: ¿Cuáles son las teorías que explican este conjunto de regularidades? Y para esa tarea se utiliza el método desarrollado en el primer capítulo.

- El tercer capítulo presenta las tres principales escuelas que conforman la actual ciencia económica: neoclásica, clásica y keynesiana. Aquí se les denomina «teoría estándar». La conclusión es que las teorías estándar pueden explicar algunas de las regularidades establecidas en el segundo capítulo, pero no pueden explicar *todas*.

- Frente a esas dificultades, surge la necesidad de desarrollar nuevas teorías. En efecto, nuevas escuelas económicas se han desarrollado en los últimos años. El libro presentará una de ellas, que podría denominarse la escuela de la economía política de la inclusión y la exclusión. Las teorías de esta escuela se presentarán primero como *teorías parciales* y luego como una *teoría unificada* del capitalismo. El cuarto capítulo presenta tres teorías parciales del capitalismo, en el sentido que buscan explicar los tres tipos de capitalismo en los que se puede dividir el sistema capitalista: el primer mundo y el tercer mundo, que a su vez se divide en dos grupos, según su historia colonial. Estas teorías parciales logran explicar cinco de las siete regularidades establecidas en el segundo capítulo.

- La explicación de las dos regularidades restantes requiere de una teoría unificada del capitalismo. El primer mundo no solo es más rico que el tercer mundo, sino que es más igualitario, y estas dos diferencias se mantienen en el tiempo. ¿Por qué persisten estas diferencias? En realidad, estas regularidades indican que existe una desigualdad de ingresos *dentro* de los países y también *entre* países, y que ambas desigualdades persisten en el tiempo.

- Ciertamente, estas desigualdades constituyen una paradoja, pues ocurren en los periodos de mayor integración mundial y constituyen, que duda cabe, el problema social fundamental de nuestra época. Por ejemplo, la violencia social en que vivimos tiene su origen en estas desigualdades, como se mostrará en este libro. Estas desigualdades son, por otro lado, las regularidades más desafiantes de comprender desde el punto de vista científico. Los capítulos 5 al 8 están dirigidos a explicar estas dos regularidades trascendentales. Los capítulos 5 al 7 construyen los elementos de la teoría unificada. El octavo capítulo muestra la teoría unificada del capitalismo, que es una teoría que intenta explicar —a partir de las tres teorías parciales que logran explicar los tres tipos de

capitalismo, pero tomados separadamente— el funcionamiento del sistema capitalista como un todo.

- ¿Cómo se llega de la ciencia positiva, que busca explicar cómo es el mundo, a la ciencia normativa, que busca establecer cómo debería ser el mundo? ¿Quiénes son los actores sociales que pueden cambiar el mundo social en que vivimos? A pesar de los avances en la ciencia económica, como indican las distinciones del Premio Nobel, los problemas sociales del mundo en que vivimos tienden a mantenerse. Ciertamente, muchas cosas han cambiado en la realidad social; pero otras no han cambiado, como es el caso del grado de desigualdad dentro y entre países, que podríamos llamar la *desigualdad global*. La desigualdad global ha persistido en un ambiente de gobiernos crecientemente democráticos. ¿Por qué? El noveno capítulo responde esta pregunta integrando las teorías positivas y normativas de la ciencia económica.

- La pregunta fundamental del libro es entender en qué mundo social vivimos y cómo la ciencia económica sirve para esa tarea. Las respuestas a las que llega el libro se presentan en las conclusiones.

- ¿En qué se parecen y en qué se diferencian los conocimientos que produce la ciencia económica de aquellos que producen las ciencias naturales? Se trata de comparar la naturaleza del conocimiento científico cuando el objeto de estudio es el mundo físico, biológico o social. Estas diferencias se presentan en el apéndice. Confinar a un apéndice esta pregunta no debe entenderse como que su importancia es mínima; se debe, más bien, a que los resultados del libro se pueden entender sin entrar al estudio de esas diferencias. Naturalmente, amigo lector, entender las diferencias entre las ciencias naturales y la ciencia económica le enriquecerá su bagaje científico, y le permitirá comprender más la naturaleza de la ciencia económica y la validez de las conclusiones del libro.

- El lenguaje constituye una de las grandes dificultades para el desarrollo de las ciencias. Se necesita utilizar un lenguaje preciso para expresar ideas precisas; y un lenguaje abstracto para expresar ideas complejas. Pero este lenguaje científico tiene el efecto de hacer poco accesible la ciencia al público general. En economía también se da este dilema. Para ayudarlo a resolver este problema, amigo lector, se acompaña un glosario de los principales términos económicos utilizados en el texto.

Este pequeño libro se enfrenta entonces al desafío de tratar grandes temas en pocas páginas y de manera accesible. Para lograrlo se ha tenido que sacrificar varias cosas. Una de ellas es la omisión a la amplia literatura teórica de la economía y a todos los debates que en ella existen. Este libro no intenta ser ni un tratado de economía ni una historia del pensamiento económico.

El libro se ha construido sobre la base de un alto nivel de abstracción de la realidad. Para simplificar la realidad y hacerla más comprensible, supondremos que la economía produce un solo bien, con lo cual haremos abstracción del papel que juegan en la producción y distribución los distintos sectores de la economía real, tales como agricultura, manufactura, minería y petróleo. También se supondrá sociedades abstractas que funcionan con pocos mercados. Pero el beneficio es que el funcionamiento de estas sociedades abstractas será muy fácil de comprender; es decir, el mundo real y complejo en que vivimos será posible de ser comprendido en sus aspectos esenciales con estas sociedades abstractas.

Pablo Picasso, genio de la pintura abstracta, dijo en una oportunidad:

> El arte no es la verdad. El arte es una mentira que nos permite acercarnos a la verdad, al menos a la verdad que nos es comprensible. El artista debe descubrir cómo convence al público de toda la veracidad de sus mentiras.

Sustituya, amigo lector, «la ciencia» en lugar de «el arte», luego «el científico» en lugar de «el artista» y también «teoría» en lugar de «mentira» en esta cita, vuélvala a leer y tendrá un buen resumen del objetivo que

persigue este libro. Solo hay que notar que en la ciencia, abstracción significa simplificación.

Utilidad práctica del libro

¿Para qué le serviría a usted, amable lector, el conocimiento sobre el funcionamiento de la sociedad en que vivimos? Ciertamente, tendrá una utilidad personal, como sería la de ayudarle a entender mejor el mundo donde vive y tomar mejores decisiones sobre el uso de sus propios recursos. Por ejemplo, podrá saber dónde colocar su dinero, pues sabrá cómo lo hacen otros.

Pero tenga en cuenta que la ciencia económica no puede decirnos cómo deberíamos utilizar nuestros recursos sino, por el contrario, la ciencia económica busca explicar cómo nosotros decidimos el uso de nuestros recursos y cuáles son las consecuencias de este comportamiento individual sobre el agregado en la sociedad. Según como usemos nuestro dinero, podríamos estar creando, por ejemplo, más desempleo o menos desempleo en el agregado.

La mayor utilidad de esta obra será en la acción colectiva, pero no por eso es menos importante. En general, el conocimiento científico de la realidad nos ayuda a elevar nuestro nivel de conciencia sobre la naturaleza de los problemas del mundo. Pero, en el caso de la ciencia económica, existe un efecto adicional y muy importante. Bajo las condiciones actuales en que funciona el mundo social, tenemos poder individual para decidir sobre algunas cosas pero no sobre otras. El mayor conocimiento de la realidad nos puede sugerir formas novedosas de acción individual, como se mencionó anteriormente, y también de acción colectiva para poder decidir sobre más cosas que nos afectan. Así, tendremos instrumentos para cambiar el mundo social en que vivimos, en lugar de esperar que los problemas sociales se resuelvan por las acciones de las elites políticas y económicas.

Luego de leer este libro, usted, amigo lector, estará probablemente en condiciones de comprender mejor el funcionamiento de la sociedad en que vive, pues su comprensión estará basada en el conocimiento científico. Podrá, en consecuencia, manejarse mejor en este mundo social. Existe un principio fundamental del conocimiento científico que dice: «No hay cosa más práctica que una *buena* teoría». El subrayado es fundamental, pues el principio no se refiere a cualquier teoría. Esta es la utilidad de la ciencia.

En un mundo donde la distribución del poder económico es muy desigual, el conocimiento científico de la realidad social es una herramienta de redistribución de ese poder. Para ello, necesitamos entender no solo nuestro mundo local, sino también el mundo global. Upton Sinclair, un escritor estadounidense, dijo en 1907: «Es difícil conseguir que una persona entienda algo cuando su salario depende de algo que ella no entiende». El funcionamiento del mercado del trabajo es, en efecto, un tema central en este libro.

En suma, el libro está dirigido a cualquier persona interesada en entender el mundo social en que vivimos. Este requisito es quizá el único importante para su lectura.

Capítulo 1

Fundamentos de la ciencia económica

¿Qué cosa es el conocimiento científico?, ¿cómo se llega al conocimiento científico? Son preguntas que nos han atormentado y nos han desafiado desde los orígenes de la civilización y seguirán haciéndolo en el futuro. No existen respuestas sencillas. Una rama de la filosofía, llamada indistintamente filosofía de la ciencia, teoría del conocimiento, metodología o epistemología, se ocupa de estudiar este tema. El término epistemología significa literalmente lógica del conocimiento científico (del griego *episteme*, conocimiento).

Se puede decir que un conocimiento es científico cuando está libre de error. Luego, para llegar al conocimiento científico debemos de seguir un conjunto de reglas o principios que nos proteja del riesgo de entender algo de manera errada. Una forma es que esas reglas constituyan un sistema lógico. Pero, ¿cómo se pueden establecen estas reglas? Estas reglas no podrían ser dadas por otra ciencia. Si una ciencia partiera de otra ciencia, ¿de dónde partiría esta segunda ciencia? Tendría que partir de una tercera, la cual a su vez necesitaría de otra y así sucesivamente; por lo tanto, en este intento terminaríamos en el problema lógico de la regresión continua. Se necesita un punto inicial de apoyo para avanzar. Las reglas del conocimiento científico solo pueden provenir de fuera de la ciencia: la epistemología es ese punto de apoyo.

En los países anglófonos, a una persona que culmina sus estudios de doctorado en Economía, el diploma que le entrega la universidad tiene

la inscripción *Philosophy Doctor in Economics,* o *Ph.D. in economics*, es decir Doctor de Filosofía en la Ciencia Económica. Este término hace referencia a que la persona está versada en epistemología, esa rama de la filosofía que se ocupa de la lógica del conocimiento. Igual se aplica a los doctorados en otras ciencias.

Para hablar de ciencia, entonces, una breve introducción a la epistemología se hace necesaria. No se trata de presentar aquí todo el amplio debate que existe sobre el tema, sino de determinar la lógica del conocimiento científico, y las reglas que de ella se derivan, que se utilizará en este libro. Las conclusiones, en cada uno de sus capítulos, serán entonces transparentes, pues ellas serán obtenidas utilizando estas reglas. Así, el libro podrá mostrar a la ciencia económica en acción.

La epistemología popperiana

Las ciencias buscan establecer relaciones entre objetos. Estos objetos pueden ser mentales o fácticos. En el primer caso, las relaciones se dan entre ideas y dan lugar a las ciencias formales —como la matemática y la lógica—. En el segundo caso, las relaciones se establecen entre objetos del mundo real y dan lugar a las ciencias fácticas, como las ciencias naturales y las ciencias sociales.

Para llegar al conocimiento en una ciencia formal, las relaciones que se establecen deben estar libres de contradicciones lógicas internas. Esta es una condición necesaria y suficiente para no caer en error. En las ciencias fácticas, en cambio, ese criterio constituye solo una condición necesaria, pues en este caso la realidad cuenta.

¿Cuáles deberían ser los principios epistemológicos en los que se sustente una ciencia fáctica? En este libro adoptaremos el principio que propuso el filósofo Karl Popper. Las relaciones que establecen las ciencias fácticas toman la forma de enunciados o proposiciones. Una ciencia fáctica se compone de proposiciones que deben cumplir con dos reglas: (a) deben ser lógicamente correctas; (b) deben ser

empíricamente falsables o refutables. Así, Popper introdujo el principio de la falsabilidad como la regla básica del conocimiento en la ciencia, como la línea divisoria entre ciencia y no ciencia. Veamos algunos ejemplos. La siguiente proposición falla con respecto a la primera regla:

La tierra es un plato que reposa en el
caparazón de una tortuga gigante

Esta proposición no constituye un argumento lógico si no se nos dice sobre qué descansa la tortuga. Siempre se cuenta la graciosa solución que alguien pretendió darle a este problema diciendo: «descansa sobre otra tortuga pues». Y la siguiente proposición falla con respecto a la segunda regla:

Un chamán le dice a su cliente:
«si tienes fe, mi medicina te sanará»

Si el cliente no se sana, el chamán le dirá que no tuvo fe. Como la fe no es observable, el cliente no tendrá manera de probarle a este chamán que su medicina es inútil. Esta proposición es, en realidad, una tautología, pues siempre será verdadera, no hay manera de refutarla.

La economía es una ciencia social porque busca conocer el funcionamiento de las sociedades humanas.[1] Las sociedades humanas son realidades complejas. Se puede decir que una realidad es compleja cuando los elementos que la constituyen son heterogéneos y numerosos, y cuando los factores que afectan los resultados de las interacciones entre los elementos son múltiples. Luego, el mundo social es una realidad

[1] Una nota de cautela. En castellano, la palabra «economía» tiene dos acepciones, esto es, se puede referir a la ciencia o a la realidad que la ciencia justamente intenta estudiar, como en «economía peruana». Para evitar estas confusiones, aquí se tratará de usar «ciencia económica» para referirse a la primera acepción y para la segunda siempre irá con un adjetivo. En inglés la diferencia es clara: *economics* para la primera y *economy* para la segunda.

compleja, pues las personas, que son las unidades que conforman el mundo social, son muy heterogéneas y numerosas, y los factores que afectan los resultados de las relaciones interpersonales son múltiples.

Cuando una realidad es compleja, no todo aspecto de ella puede ser objeto de conocimiento científico, sino únicamente aquellos aspectos que muestren relaciones sistemáticas. Tales relaciones existirán si existe *repitencia* en las relaciones; es decir, si es posible representar la realidad en la forma de un proceso. Para ser comprendidas, por lo tanto, las realidades complejas deben ser transformadas a un proceso.

¿Qué es un proceso? Es una serie de actividades, de una duración determinada, que tiene un resultado y que se repite periodo tras periodo. En un proceso existen elementos del exterior que ingresan y elementos que salen del interior del proceso. A los primeros los llamaremos *elementos exógenos* y a los segundos *elementos endógenos*. La figura 1.1 ilustra el concepto de proceso, en donde el segmento t_0-t_1 representa la duración del mismo, X representa el conjunto de elementos exógenos, e Y el conjunto de elementos endógenos, que constituyen el resultado o producto del proceso.

En un proceso también existe un mecanismo por el cual los elementos exógenos producen los elementos endógenos. Pero lo que sucede al interior del proceso, el mecanismo de interacción entre los elementos exógenos y endógenos no es observable (esto se indica en la figura 1.1 por medio del área sombreada.) Si fuera observable, el interior del proceso sería un proceso en sí mismo, con elementos endógenos y

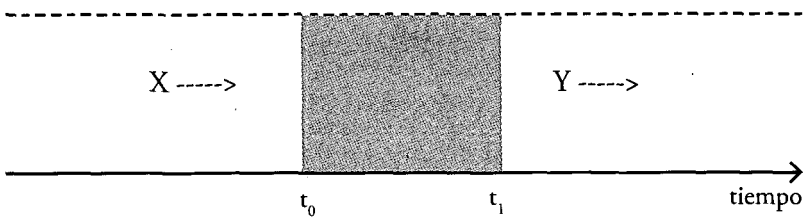

Figura 1.1 Representación gráfica de un proceso

exógenos y un interior observable, el cual a su vez sería otro proceso, y así sucesivamente; en consecuencia, terminaríamos en el problema lógico de la regresión continua. En última instancia, debe existir algo escondido detrás de las cosas que observamos. La ciencia busca, precisamente, descubrir estos factores subyacentes, el por qué de los hechos que observamos.

Considere el siguiente ejemplo: la producción de un bien, como la papa, puede ser transformada en un proceso, que llamaremos el proceso productivo. Los elementos exógenos X se refieren al conjunto de factores que ingresan en la actividad de producir papa, tales como cantidades de semilla de papa, tierra, agua, trabajadores y fertilizantes; los elementos endógenos Y se refieren a las cantidades producidas de papa y a los deshechos. La caja sombreada de la figura 1.1 se refiere al conocimiento tecnológico de los agricultores, a la fisiología de la planta y a todos los mecanismos biológicos que transforman la semilla en el producto. Si se introduce al proceso productivo la misma cantidad de elementos X se obtendrá —ignorando el efecto aleatorio del clima— la misma cantidad producida de papa, año tras año.

Dado que un proceso se repite periodo tras periodo, las relaciones entre los elementos endógenos y exógenos pueden ser continuamente observadas, lo que da lugar al surgimiento de relaciones sistemáticas o de regularidades empíricas. La ocurrencia de regularidades es condición necesaria para que una realidad sea objeto de investigación científica. Un mundo caótico, en el que las regularidades están ausentes, presenta dificultades serias para ser objeto de investigación científica con la epistemología popperiana; del mundo caótico no se va a ocupar este libro.

Un listado de todos los elementos endógenos y exógenos de un proceso incluiría elementos observables y no observables. Llamemos *variables endógenas* y *variables exógenas* a los elementos observables. Con esta denominación se quiere decir que es posible medir las variaciones de estos elementos del proceso.

Ciertamente, presentar el listado completo de todos los elementos que participan en un proceso sería como presentar la realidad en un

mapa a la escala 1:1. Para su comprensión, la realidad compleja tiene que ser reducida a un proceso simple. Pero eso significa dejar fuera del estudio varios elementos del proceso. Así, llegamos a construir un proceso abstracto, en el cual solo se incluirán algunos elementos del proceso, donde el criterio de inclusión es que sean los más importantes. Al método de tomar en cuenta solo algunos elementos de la realidad —los que son los más importantes—, e ignorar el resto, se le llama *abstracción*.

¿Cómo se decide cuáles son los elementos importantes en un proceso?, ¿cómo se construye un mundo abstracto? Utilizamos supuestos. El método de la abstracción implica introducir supuestos sobre lo que es importante y lo que se puede ignorar. Este conjunto de supuestos constituye una *teoría*. La teoría tiene un objetivo: explicar las relaciones del proceso. Por lo tanto, la teoría determina cuáles son las variables endógenas del proceso, aquellas que serán el objeto de estudio; también determina las variables exógenas, aquellas que considera como los factores más importantes que afectan a las variables endógenas; finalmente, la teoría también hace supuestos sobre los mecanismos de interacción entre las variables endógenas y exógenas.

¿Se podría utilizar la observación empírica misma para establecer los supuestos de la teoría? No. La razón es simple: la teoría busca determinar precisamente los factores que subyacen a esta observación empírica. Por lo tanto, la idea de que los supuestos deben ser realistas —afincados en la realidad— no tiene ningún sentido. Los supuestos de una teoría no necesitan justificación empírica, ni de ninguna otra índole. Si los supuestos de una teoría tuvieran que justificarse, se necesitaría otro conjunto de supuestos que la justifiquen, y estos últimos supuestos a su vez necesitarían de otros, y así sucesivamente; por tanto, terminaríamos en el problema lógico de la regresión continua.

Los supuestos de una teoría se establecen, en consecuencia, de manera arbitraria, como si fuesen axiomas, sin justificación. No existe otra forma de elegirlos. Esto puede parecer poco o nada científico. Pero los supuestos se eligen como parte de un mecanismo de tanteo, de prueba

y error, para llegar al mejor conjunto de supuestos, es decir, a la buena teoría, a la buena representación abstracta de la realidad. Por haber sido elegidos arbitrariamente, los supuestos de una teoría necesitan confrontación con la realidad para determinar si el mundo abstracto y simple que la teoría ha construido es una buena aproximación del mundo real y complejo que intenta representar. Si los supuestos de una teoría fallan, se abandona la teoría y se le reemplaza por otra teoría. No existe criterio a priori para demostrar que una teoría es superior a otra; la prueba es empírica.

Teoría es entonces un artificio lógico para reducir una realidad compleja a un mundo abstracto y simple. Este mundo abstracto será mucho más comprensible. Considere nuevamente el ejemplo del proceso de producción de papa. Si el proceso de la figura 1.1 mostrara un listado de todos los elementos que intervienen tendríamos una descripción completa de la realidad —¡una lista enorme!—, pero sería incomprensible; si se incluyera en el listado solo algunos elementos, aquellos que se suponen son los más importantes, tendríamos un mundo abstracto pero más simple de ser comprendido.

La teoría así establecida puede, en principio, fallar porque ha eliminado elementos del proceso; pero, si pudiendo fallar, se muestra como una buena aproximación de la realidad, es una buena teoría. Para determinar esta conformidad, la teoría tiene que cumplir con los dos requisitos de la metodología popperiana señalados anteriormente: tiene que ser lógicamente correcta y empíricamente falsable o refutable. Como una metodología es un conjunto de métodos, a continuación se presentará un método para operar la metodología popperiana.

El método alfa-beta

La metodología popperiana sostiene que las proposiciones de una teoría tienen que ser empíricamente refutables o falsables para ser científicas. Pero, ¿cómo se puede ejecutar este principio de falsabilidad?

Tenemos que construir un método particular que sea consistente con el principio.

El conjunto de proposiciones de una ciencia se puede separar en dos categorías: las primarias y las derivadas, siendo las segundas lógicamente derivadas de las primeras, y ninguna proposición primaria puede ser derivada lógicamente de otra proposición primaria. Llamemos *proposiciones alfa* a las primarias y *proposiciones beta* a las derivadas.

El conjunto de supuestos de una teoría se puede denominar ahora como el conjunto de proposiciones alfa, pues es el conjunto de supuestos primarios o proposiciones primarias. Estos supuestos, como vimos antes, se construyen como axiomas, pues no necesitan justificación. Las proposiciones alfa hacen supuestos sobre los factores esenciales del proceso y sobre las fuerzas y motivaciones que subyacen a las interrelaciones entre las variables endógenas y exógenas. Las proposiciones alfa son no observables porque se refieren a los factores que subyacen en los hechos observados.

Las relaciones que son observables son las que se dan entre las variables endógenas y exógenas de la teoría. Si estas relaciones son derivadas lógicamente de las proposiciones alfa, ellas constituirán el conjunto de proposiciones beta. Por lo tanto, las proposiciones beta indican las implicancias o predicciones empíricas de la teoría. Si la teoría es una buena aproximación de la realidad se debería observar en el mundo real lo que sus proporciones beta dicen; si la teoría no es buena, sus proposiciones beta no deben coincidir con nuestras observaciones del mundo real. Aunque el conjunto de proposiciones alfa se establece de manera arbitraria, el conjunto de proposiciones beta ya no puede ser arbitrario, pues predice lo que debemos observar en el mundo real.

Una proposición beta es, en consecuencia, una proposición empíricamente refutable; esto es, que en principio puede ser empíricamente falsa. Una buena teoría es aquella que pudiendo ser empíricamente falsa, no lo es. Una buena teoría es como una persona honrada, que es aquella que nunca ha delinquido, pudiendo hacerlo; si nunca tuvo la oportunidad de delinquir, no se puede decir que es honrada. Si una teoría

da lugar a proposiciones beta que nunca pueden fallar empíricamente, esta teoría es una tautología, inútil para el conocimiento científico. Las proposiciones beta, por construcción, pueden ser empíricamente falsas, es decir, pueden ser refutadas por los hechos del mundo real.

La teoría es entonces un conocimiento a priori, un artificio lógico que nos permite construir un mundo abstracto que sea una buena representación del mundo real. Si no hay teoría, no puede haber conocimiento científico del mundo real. Pero la teoría necesita confirmación empírica. El conocimiento a priori necesita una confirmación a posteriori. ¿Por qué? Porque los supuestos fueron establecidos arbitrariamente. Si en esta confrontación teoría y realidad son inconsistentes, es la teoría la que pierde —¡no la realidad!—; esto es, queda demostrado que la selección arbitraria de sus supuestos fue errónea. Aun cuando una teoría es un sistema lógicamente correcto, puede ser empíricamente falsa.

Así, la confrontación empírica de una teoría puede realizarse solo de forma indirecta, a través de las proposiciones beta. Son estas predicciones empíricas de la teoría las que pueden o no coincidir con la realidad. Cuando esta consistencia se da, se dice que la teoría explica la realidad. El mundo abstracto de la teoría se parece al mundo real. El mundo real se puede comprender, en lo esencial, a partir del mundo abstracto que ha construido la teoría. En caso contrario, la teoría es empíricamente falsa.

Considere el siguiente ejemplo. Una teoría sostiene que «la figura F (no observada) es un cuadrado». Luego, una proposición beta de esta teoría diría que «las dos diagonales deben ser iguales». Si existe información sobre las diagonales y se muestra que no son iguales, se puede concluir que la teoría es falsa; si se muestra que son iguales, la evidencia empírica es consistente con la teoría.

A causa de su papel empírico, las proposiciones beta incluyen relaciones entre variables exógenas y endógenas solamente, que son observables. ¿Por qué? Porque si se incluyeran variables inobservables, las proposiciones beta no serían refutables. Eligiendo cambios apropiados en los elementos exógenos inobservables *siempre* se podría hacer

coincidir la teoría con la realidad. Al introducir elementos exógenos inobservables, una teoría se vuelve inmortal. Así, por ejemplo, si teoría y realidad no coinciden se puede atribuir esa diferencia a un cambio en las expectativas de la gente. Pero como las expectativas no son observables, este elemento se vuelve un comodín, y la teoría ya no es refutable. Estamos frente al problema del chamán antes mencionado.

Como las proposiciones beta predicen una relación particular y observable entre las variables endógenas y exógenas de una teoría, también se puede decir que predicen relaciones de causa y efecto, que denominamos *relaciones de causalidad.* Las variables exógenas tienen un efecto en una dirección particular —positiva o negativa— sobre las variables endógenas. Debe quedar en claro que las relaciones de causalidad se derivan de una teoría.

Sobre el proceso de falsación, dado que la teoría construye un mundo abstracto, donde solo algunos factores del proceso son tomados en cuenta, los datos de la realidad difícilmente pueden calzar a la perfección con las proposiciones beta. Por lo tanto, la consistencia empírica entre teoría y realidad solo puede darse en términos estadísticos: debe ser válida en la generalidad de los casos, en el promedio, y ciertamente pueden existir excepciones. Así, se requiere del análisis estadístico para aceptar o rechazar una proposición beta.

Encontrar una sola observación en la que una proposición beta falle es insuficiente para rechazar la teoría, en tanto el valor estadístico de una observación es nulo. Esa observación podría responder a una excepción, a una desviación de la norma. El test debe ser, entonces, estadístico. Por ejemplo, si una teoría predice que fumar causa cáncer, encontrar una persona que fuma pero no tiene cáncer no puede invalidar la teoría. Un contra ejemplo sirve para destruir un teorema en matemáticas, pero no puede destruir una teoría en las ciencias fácticas. Es la relación entre los valores promedios de las variables lo que cuenta. En consecuencia, debe establecerse una distinción entre un *error estadístico de la teoría,* que se refiere a la existencia de excepciones, y una *falla de la teoría,* que se refiere a una falla sistemática.

Denominaremos el *método alfa-beta* al uso de las proposiciones alfa y beta para llegar a construir proposiciones lógicamente correctas y empíricamente refutables. Este método es, por lo tanto, una forma de operar la metodología popperiana. La lógica del método alfa-beta es internamente consistente. Una realidad compleja es susceptible de conocimiento científico cuando, por medio de una teoría, es representada como un proceso abstracto, donde existe repetición y las regularidades empíricas pueden ocurrir, y estas regularidades son utilizadas para someter a las proposiciones beta a la refutación estadística.

Si las proposiciones beta se ajustan bien a los datos empíricos, puede decirse que la teoría es *consistente* con la realidad. No podemos decir que la teoría es *verdadera*. Esto es así porque las mismas proposiciones beta podrían derivarse de otra teoría. No hay una relación de uno a uno entre las proposiciones alfa y las proposiciones beta. Por tanto, la aseveración acerca de que una teoría «explica» la realidad tiene un significado muy preciso: sus proposiciones beta son consistentes con los datos empíricos. Si los datos empíricos no coinciden con las proposiciones beta, la teoría falla. En este caso, se formula una nueva teoría y así, por prueba y error, se llegará a establecer una buena teoría.

El método alfa-beta se representa en la figura 1.2. A partir de un conjunto de supuestos primarios α_1, se deriva lógicamente el conjunto de proposiciones empíricamente observables β_1 (indicado mediante la flecha doble), el que debe ser sometido a la refutación estadística (indicado mediante la flecha simple). La flecha doble indica el procedimiento lógico; la flecha simple señala el procedimiento operacional, la actividad a realizar. El análisis estadístico (señalado mediante el símbolo $\approx$) implica la búsqueda de una conformidad estadística entre las proposiciones beta que predicen relaciones particulares que deben existir entre las variables endógenas y exógenas y el conjunto de relaciones, asociaciones o correlaciones empíricas entre las variables endógenas y exógenas, que denominaremos b. Si se muestra que estadísticamente $\beta_1 = b$, entonces α_1 es consistente con la realidad. Si $\beta_1 \neq b$, entonces α_1 es falso y debe desarrollarse una nueva teoría α_2, y así continuar con el algoritmo hasta obtener una buena teoría.

$\alpha_1 \Rightarrow \beta_1 \rightarrow [\beta_1 \approx b]$

Si $\beta_1 = b$, α_1 es consistente con b y explica la realidad
Si $\beta_1 \neq b$, α_1 es falso y no explica la realidad. Entonces:

$\alpha_2 \Rightarrow \beta_2 \rightarrow [\beta_2 \approx b]$

Si... (el algoritmo continúa)

Figura 1.2. El método alfa-beta

Las relaciones o asociaciones empíricas (indicadas por la letra b en la figura 1.2) se calculan a partir de una base de datos y a la cual se le aplica una teoría estadística de inferencia. Por lo tanto, el conjunto de asociaciones empíricas obtenidas cambiará cuando se cuente con una nueva base de datos o con nuevas teorías estadísticas. Innovaciones en los instrumentos de medición también modificarán la base de datos. Estos casos de modificaciones darán lugar a nuevas pruebas estadísticas sobre la validez empírica de la teoría.

Dado el tamaño del conjunto de datos de la realidad, indicado por la letra b, una teoría será falsa si se encuentra que una proposición beta es estadísticamente falsa. Es decir, es suficiente que la teoría falle en una sola de sus proposiciones beta. En el ejemplo utilizado anteriormente es suficiente mostrar que estadísticamente las diagonales no son iguales para concluir que la figura F no es un cuadrado. Así, una teoría es empíricamente consistente con la realidad si y solo si ninguna de sus proposiciones beta es estadísticamente falsa.

Cuando se confrontan empíricamente varias teorías a la vez, dada una base de datos, algunas de ellas resultarán falsas y otras resultarán consistentes. Con nueva información, algunas de las teorías del último grupo serán falsas y otras seguirán siendo consistentes, y así sucesivamente. Aquellas que sobrevivan a todos los procesos de confrontación

serán las teorías *buenas*. Y reinarán hasta que una nueva información disponible conduzca a la aparición de inconsistencias.

Un ejemplo sencillo puede ilustrar el método alfa-beta. Suponga una figura F que no es observable por el investigador. Considere la siguiente teoría 1: «la figura F es un cuadrado». Si la figura es un cuadrado, entonces sus dos diagonales deberían ser iguales —una proposición beta—. Si estadísticamente las dos diagonales no son iguales, se puede concluir que F no es un cuadrado y la teoría es falsa. Si las dos diagonales son iguales, ¿se puede concluir que la teoría es verdadera? No. No se puede concluir que F es un cuadrado y que la teoría es verdadera, debido a que esta evidencia es también consistente con, digamos, la teoría 2: «la figura F es un rectángulo». Solo se puede decir que la teoría 1 es consistente con la información. Las teorías 1 y 2 son consistentes con la evidencia empírica que muestra que las diagonales son iguales. En este caso, el antónimo de «falso» no es «verdadero» sino «consistente».

El método alfa-beta es consistente con el principio de falsación de Popper. Por lo tanto, el método alfa-beta se inscribe en la lógica de la epistemología popperiana, pues nos permite refutar teorías —proposiciones alfa— sobre la base de la derivación lógica de las predicciones empíricas de la teoría que son refutables —proposiciones beta— y su confrontación con los datos de la realidad. Las proposiciones beta nos dicen qué cosa hay que observar para refutar una teoría. Este procedimiento elimina cualquier posibilidad de estratagemas inmunizadoras —como utilizar elementos no observables en la explicación— que busquen salvaguardar a la teoría de la refutación empírica.

La metodología alfa-beta implica que una teoría está sujeta a continuas confrontaciones con la realidad. La buena teoría es aquella que ha sobrevivido a numerosos intentos de refutación. Pero la confrontación será continua, pues nuevos datos pueden refutarla o la aparición de nuevas teorías pueden superarla. Aceptar teorías es siempre una decisión provisional, pero su rechazo es definitivo. A este procedimiento se le puede denominar el *principio de la razón insuficiente*: dada la información que existe, no hay razón suficiente para refutar la teoría. Este

procedimiento es similar al que se utiliza en las pruebas estadísticas, donde rechazar la hipótesis nula es definitivo, pero aceptarla es provisional, en el sentido que, dada la información existente, no existe razón alguna para rechazarla.

El método alfa-beta también nos muestra que los datos por *sí* mismos no pueden explicar el mundo real. La razón es simple: los datos empíricos por sí mismos no pueden producir relaciones de causalidad. Los datos pueden producir *correlación* estadística, es decir, una asociación entre los datos observados que indica que se mueven de manera conectada, sea en la misma dirección —correlación positiva—, o en dirección contraria, cuando uno aumenta el otro disminuye —correlación negativa—. Pero esto no implica *causalidad.* Por ejemplo, no se puede decir que los bomberos son causa de los incendios porque observamos que donde hay un incendio hay usualmente bomberos. No existe una ruta lógica desde la correlación hacia la causalidad. Como hemos visto aquí, la causalidad requiere una teoría, pues la causalidad está contenida en las proposiciones beta, las cuales se derivan de una teoría.

El método alfa-beta en la ciencia económica

El campo y método de la ciencia económica puede ser definida de manera más precisa con la ayuda del concepto de proceso. La figura 1.1 nos servirá de referencia. La economía estudia un proceso particular, el proceso económico. La producción y distribución constituyen las variables endógenas, el objeto de estudio. Las variables exógenas y los mecanismos de interacción entre estas y las variables endógenas se establecen mediante un conjunto de supuestos, una teoría económica. Esta teoría reduce el proceso económico del mundo real a otro proceso de un mundo abstracto pero más simple.

En términos del método alfa-beta, una teoría económica corresponde a las proposiciones alfa, de la cual se derivan las proposiciones beta. Una proposición beta muestra la relación entre las variables endógenas

y exógenas. En consecuencia, para cada variable endógena de la teoría existirá una proposición beta; así, habrá tantas proposiciones beta como variables endógenas contenga la teoría. El uso del método alfa-beta en la ciencia económica tendrá sus particularidades, que pasamos a señalar.

Una teoría económica es una familia de modelos

En la ciencia económica puede ocurrir que las proposiciones beta no puedan ser derivadas directamente de las proposiciones alfa, debido a que estas son proposiciones muy generales. En este caso, se requiere transformar las proposiciones alfa en un conjunto de proposiciones operacionales. Este nuevo conjunto de proposiciones se llama el *modelo* de la teoría. Un modelo teórico es un conjunto de supuestos que incluye las proposiciones alfa y un subconjunto de *supuestos auxiliares consistentes*. La consistencia se refiere a que los supuestos auxiliares no pueden contradecir los supuestos primarios de la teoría. Una teoría es, entonces, una familia de modelos, donde el conjunto de proposiciones alfa constituye el núcleo de toda la familia. La lógica de la construcción de modelos no es proteger la teoría, sino por el contrario es hacerla efectivamente falsable.

En la figura 1.2, considere que α_1 y α_2 sean dos modelos de la teoría α. El algoritmo de falsación es aplicado entonces a estos modelos. Si el primer modelo no pasa la prueba, el segundo ingresa a la prueba, y así sucesivamente. Para que una teoría sea falsa, todos los modelos de la familia deben fallar. El algoritmo requiere, por tanto, que el número de modelos que pueda generar la teoría sea finito. Una buena teoría es aquella que da lugar a un número limitado de supuestos auxiliares consistentes y, por lo tanto, de modelos; en caso contrario, la teoría no será refutable. Llevada al proceso de falsación, si todos los modelos fallan, la teoría es empíricamente falsa, y se hace necesario construir una nueva teoría.

En términos de teoría y modelo, podemos decir que la teoría popperiana del conocimiento nos brinda los principios para llegar al

conocimiento científico, una lógica del conocimiento científico. Para ser operacional, esta teoría del conocimiento debe ser transformada en un modelo. A partir de este se puede generar un conjunto de reglas prácticas para aceptar o rechazar teorías económicas. El método alfa-beta es un modelo particular de la teoría popperiana del conocimiento.

En la ciencia económica, una teoría económica se somete a la falsación empírica a través de un modelo. Si el modelo no es refutado por los hechos de la realidad, la teoría se acepta; si el modelo falla, se debe construir otro modelo de la teoría y someterlo a la prueba de la falsación. Si todos los modelos de la teoría fallan, entonces la teoría falla. Este método requiere que la teoría tenga un número finito de modelos. Si el número de modelos es infinito, la teoría no es falsable.

Sobre la utilidad práctica de la teoría económica, considérese la proposición de Popper: «Todos los organismos vivientes tienen capacidad para resolver problemas, aquellos problemas que surgen con la vida misma». ¿Cómo hacen estos organismos para resolver sus problemas? «Todos los organismos vivientes usan teorías para solucionar sus problemas», añade Popper. Los que usan teorías falsas pagan algún costo, en algunos casos el costo es su extinción como especie.

Esta proposición de Popper implica que una ameba, un hombre de la calle y un científico usan teorías para resolver problemas. Lo que distingue al científico es que su metodología para eliminar teorías falsas es más desarrollada, está elaborada conscientemente, y no está basada en instintos.[2] Por esto, la ciencia necesita una metodología explícita, impersonal, lógica y perfectible. Si realmente todos los organismos vivientes construyen sus teorías para sobrevivir y actuar, desde la ameba al científico, podría afirmarse, de nuevo, que «no hay nada más práctico que una buena teoría».

---

[2] La otra diferencia es que mientras las teorías de la ameba se construyen en el organismo de la ameba, el científico puede formular sus teorías mediante el lenguaje, de este modo puede trasladar sus teorías fuera de su organismo. Así, el científico puede estudiar a todos los organismos, inclusive a su propia especie, como en las ciencias sociales. En este caso, el científico social viene a ser el etólogo de su propia especie.

El supuesto del equilibrio económico es fundamental para derivar las proposiciones beta

¿Cómo se derivan las proposiciones beta? Un modelo debe establecer las interacciones entre los actores sociales y la solución de esas interacciones sobre las variables endógenas, dado los valores de las variables exógenas. Esta solución se denomina el *equilibrio económico*. Indica, por definición, que se trata de una situación donde ningún actor social tiene el poder y el incentivo para cambiar la situación. Y, entonces, esta situación se va a repetir periodo tras periodo, siempre y cuando las variables exógenas permanezcan fijas.

Cuando cambien las variables exógenas, la situación de equilibrio será otra, con otros valores de equilibrio de las variables endógenas. Las variables exógenas tendrán efectos sobre las variables endógenas pasando de una situación de equilibrio a otra. Así se generan las proposiciones beta. Como se puede ver, la existencia del equilibrio económico es fundamental para esta derivación.

Como hemos señalado antes, las proposiciones beta sirven para someter el modelo al proceso de falsificación. El supuesto de la existencia del equilibrio da lugar a otra implicancia empírica del modelo. Las condiciones de equilibrio también tienen implicancias empíricas, pues este equilibrio puede exigir que las variables endógenas se encuentren en ciertos rangos de valores, lo cual es observable y refutable. Luego, las condiciones de equilibrio que sean observables también serán parte de las proposiciones beta del modelo.

Las proposiciones alfa universales

La ciencia económica estudia un proceso particular: el proceso económico, que es definido como el proceso de producción de bienes y su distribución entre los distintos grupos sociales que forman las sociedades humanas. La economía busca establecer las variables exógenas que explican el resultado de la producción y distribución en sociedades

humanas particulares. Una teoría económica debe incluir entonces un conjunto de supuestos sobre las motivaciones de la gente y el mundo de las relaciones sociales en que operan, las cuales subyacen a los hechos observados en la producción y distribución en estas sociedades.

La pregunta central de la ciencia económica tiene que ver con la *ontología universalista* o particularista del proceso económico. El supuesto que adoptaremos aquí es que el proceso económico tiene particularidades para cada sociedad, pero también que existen relaciones que son universales. En otras palabras, la ciencia económica supone que existen teorías universales y teorías específicas para explicar el proceso económico en sociedades particulares. Veamos aquí las proposiciones alfa de la teoría universal y dejemos para el resto del libro el uso de las teorías específicas, que se referirán a la sociedad capitalista.

Primero, la ciencia económica supone que la naturaleza del problema económico es similar en todas las sociedades. Todas enfrentan el mismo problema de escasez, esto es, las necesidades humanas son ilimitadas pero la disponibilidad de los bienes es limitada. Como los bienes son escasos, estos deben ser producidos, pero se supone que los factores de producción son limitados. Este es el problema económico de toda sociedad. Para resolverlo, se supone que cada sociedad se organiza y establece las reglas bajo las cuales debe operar el proceso económico.

Segundo, la ciencia económica supone que los individuos (que llamaremos indistintamente agentes o actores sociales) actúan guiados por una motivación; es decir, existe una lógica que subyace a su comportamiento. A esta lógica de comportamiento se le denomina *racionalidad económica.* Este supuesto de la racionalidad lleva a los individuos a actuar de manera consistente y no errática, haciendo evaluaciones entre fines y medios. Se supone implícitamente que los individuos actúan no de manera emocional, sino de manera racional, que ellos calculan las consecuencias de sus decisiones a la luz de los objetivos que persiguen. «Comportamiento racional» no tiene significación ética, de ser buena o mala, sino que es un supuesto —una teoría— de la ciencia económica positiva sobre la existencia de una motivación que guía las acciones de los individuos.

Las proposiciones universales serán denominadas postulados, para indicar que ellas constituyen el núcleo de la ciencia económica.[3] Estos postulados son:

El postulado de la función de producción. Los bienes no se producen de la nada. Dado un estado de conocimiento tecnológico en la sociedad, la cantidad producida de cualquier bien depende de la cantidad de recursos usados como insumos.

El postulado de escasez. Las sociedades buscan resolver el problema económico: los recursos para la producción de bienes son limitados, en tanto que las necesidades humanas por estos bienes son ilimitadas. Ninguna cantidad de bienes será suficiente y siempre habrá escasez y, por lo tanto, siempre se buscará producir más cantidades de bienes.

El postulado institucional. Para resolver el problema económico, las sociedades buscan establecer un conjunto de reglas y organizaciones para llevar a cabo el proceso económico.

El postulado de racionalidad. El comportamiento individual está guiado por una motivación, la cual deber ser consistente con el contexto institucional de la sociedad en la cual el individuo vive.

El postulado de la incertidumbre. El proceso económico se desarrolla en un ambiente de incertidumbre. Los individuos no conocen con certeza las consecuencias futuras de sus acciones de hoy, ni conocen con certeza las ocurrencias de algunas variables exógenas. La gente busca, individual o colectivamente, establecer mecanismos de protección ante los riesgos de pérdidas que puedan sufrir.

---

[3] El término axioma es demasiado rígido para ser usado en la ciencia económica, donde se espera que las proposiciones jueguen un rol cambiante en la construcción de teorías. El término postulado parece ser mejor, más flexible. Supuesto es un término aun más flexible y será usado en este libro cuando se trate de teorías económicas específicas.

El postulado de las condiciones iniciales (historia). Las sociedades tienen un origen que se caracteriza por dos condiciones iniciales: sus dotaciones iniciales de recursos y el grado de desigualdad inicial en la distribución de esos recursos entre los individuos. El primero constituye una dotación inicial cuantitativa, mientras que la segunda es una dotación inicial cualitativa.

Para explicar el funcionamiento de sociedades concretas, la ciencia económica construye sociedades abstractas que se le parezcan, para lo cual utiliza los supuestos universales y también supuestos específicos sobre la sociedad bajo estudio. Ese conjunto de supuestos constituye la teoría económica del tipo de sociedad bajo análisis. Las relaciones que se establecen en esa sociedad abstracta son todas lógicamente correctas. Si las sociedades del mundo real no se parecen a la sociedad abstracta, falla la teoría; si coinciden, la teoría explica esa realidad.

La economía es una ciencia social porque estudia las sociedades humanas, como dijimos anteriormente. Ahora se puede precisar que es ciencia social en un doble sentido. Primero, porque su campo de estudio se refiere al proceso de producción de bienes y su distribución entre los grupos sociales de una sociedad.

Segundo, porque estudia el comportamiento agregado, bien sea de sociedades, bien sea de individuos. El problema es que el agregado no es la simple suma de las partes, pues existe lo que se llama en lógica la falacia de composición: lo que es cierto para el individuo no es cierto para el agregado. Un ejemplo es el comportamiento de la gente espectando un desfile: si *un* individuo se empina podrá ver mejor el desfile, pero si *todos* se empinan nadie podrá ver mejor el desfile. Un ejemplo del comportamiento de la gente en el mercado de crédito: si *un* individuo no puede devolver el crédito, él tiene un gran problema; pero si *todos* los clientes del banco no pueden devolver el crédito, es el banco el que tiene un gran problema. Una teoría económica, como sistema lógico que es, no puede caer en los problemas de falacia de composición.

Las teorías económicas son abstracciones del mundo real, donde se ignora aquellos factores que se supône no son importantes. De modo que si la teoría yerra al no poder explicar el comportamiento de una sociedad particular o el de un individuo particular, la teoría no falla. A diferencia de la matemática, donde un contra ejemplo es suficiente para destruir un teorema, en la ciencia económica un contra ejemplo empírico no refuta una teoría. La prueba tiene que ser estadística.

Debido al uso de la abstracción en la generación de una teoría, las pruebas empíricas deben ser, por eso, estadísticas. Aunque la teoría explica la generalidad de los casos, habrá algunas excepciones. Una sociedad puede ser parte de la excepción, puede ser un error estadístico. Así llegamos a la conclusión de que *pueden existir realidades sin teoría,* sea porque no existe teoría que pueda explicar al conjunto de sociedades o sea porque existe tal teoría pero existen algunas realidades que son la excepción.

Puede, asimismo, ocurrir que exista más de una teoría que explique una realidad social. La relación entre teoría y realidad no es necesariamente de uno a uno. Como sabemos, este es un problema clásico en la física. La luz, considerada una onda, actúa en algunos casos como partícula. El planeta Tierra, considerado esférico, puede algunas veces considerarse plano, sobre todo para distancias cortas. Qué aspecto de la luz —o la Tierra—, o qué faceta de la luz —o de la Tierra— se observaría con las circunstancias. En economía puede, entonces, existir más de una teoría para una misma realidad social.

Este capítulo ha mostrado los fundamentos de la ciencia económica en lo que se refiere a la epistemología. Las reglas del conocimiento científico han quedado claramente establecidas. Estas reglas serán utilizadas y respetadas a lo largo de los siguientes capítulos. Toca ahora presentar las regularidades empíricas del capitalismo, que constituye la agenda de investigación para el resto del libro.

Capítulo 2

Regularidades empíricas del capitalismo

¿Cuáles son las regularidades empíricas que sobre la producción y la distribución en el sistema capitalista se pueden derivar de los estudios empíricos que aparecen en la literatura internacional?

La pregunta puede parecer extraña. Estamos acostumbrados a este tipo de preguntas solo en el caso de la física, donde los grandes cuerpos del universo tienen trayectorias definidas. Pero en el proceso económico, donde los actores sociales son las personas, con voluntad de actuar de maneras muy distintas y bajo circunstancias diversas, ¿sería posible esperar comportamientos tan sistemáticos en el mundo social como para hablar de regularidades empíricas? La respuesta es afirmativa. Pero está sujeta a dos calificaciones, como veremos a continuación.

Naturaleza de las regularidades en el proceso económico

La primera cuestión a notar es que las «leyes económicas», que con más propiedad llamaremos aquí regularidades empíricas, se referirán solo al *agregado* del comportamiento de los individuos. Sobre el comportamiento de cada individuo, como bien se coloca la sospecha en el párrafo anterior, no podemos esperar regularidades. Cada individuo enfrenta condiciones cambiantes en su vida personal y hasta en su estado de ánimo. Sin embargo, cuando se agregan individuos, sí es posible

observar regularidades, pues las desviaciones individuales van en diferentes direcciones y tienden en el agregado a compensarse. Así, mientras unos pierden sus empleos, otros los consiguen; mientras unos pasan por un buen momento anímico, otros están pasando por uno malo.

A nivel individual podemos observar un comportamiento caótico o impredecible en la gente, pero a nivel agregado observamos regularidades. ¿No es esto contradictorio? Como se muestra en el apéndice, este problema también aparece en la física y constituye el problema teórico fundamental actual: en la teoría cuántica, que estudia el mundo de lo pequeño, el mundo subatómico, existe caos; pero en la teoría de la relatividad, que estudia el mundo de los cuerpos grandes, existe orden. En la economía, esta relación no es contradictoria, porque las desviaciones individuales tienden a compensarse en el agregado.

Como se verá más adelante, la ciencia económica construye teorías para explicar el comportamiento económico del individuo, tales como la teoría del consumidor, del productor, del inversionista. Estas teorías pueden parecer contradictorias con el objetivo de una ciencia social que busca estudiar el agregado. Pero no lo es. En realidad, y como también se mostrará más adelante, la teoría del individuo no intenta explicar el comportamiento de un individuo particular, sino el comportamiento del grupo; es decir, se construye un individuo abstracto que, se supone, es representativo de las motivaciones del grupo social bajo estudio.

Considere el siguiente ejemplo. Si el precio de la carne de res aumenta en el mercado, cada individuo reacciona de distinta manera. Unos consumirán menos cantidades, otros consumirán más cantidades, otros no modificarán la cantidad que consumen, otros siguen sin consumir y ni se enteran de la subida del precio, como los vegetarianos. Además, unos aumentan y otros disminuyen su consumo debido a cambios en su situación personal; por ejemplo, unos han conseguido empleo mientras que otros lo han perdido, unos se han casado mientras que otros se han divorciado. No podemos decir nada sobre la respuesta sistemática de cada individuo particular. Sin embargo, en el agregado todos esos cambios individuales tienden a compensarse y se

puede observar regularidades. En efecto, se observa que cada vez que sube el precio de la carne, y ningún otro precio cambia, la cantidad total consumida disminuye. En el agregado, existirá la curva de demanda y será decreciente: a mayor precio de un bien, su cantidad consumida disminuirá. Podemos esperar regularidades solo para el agregado.

Segundo, y como el ejemplo anterior lo sugiere, las regularidades empíricas son *leyes estadísticas*, válidas en la mayoría de los casos, pero no para todos los casos. Las regularidades en economía no son leyes determinísticas, que son válidas *siempre*, como la ley de la gravedad en física.

Dos tipos de capitalismo: primer mundo y tercer mundo

Para establecer las regularidades empíricas, definiremos como sociedades capitalistas a aquellos países que han venido operando por largo tiempo bajo las reglas y organizaciones del capitalismo —como son la propiedad privada del capital, el sistema de mercado como forma de intercambio de bienes y, principalmente, la existencia del mercado laboral—. El conjunto de países capitalistas será dividido en dos grupos: los del *primer mundo* y los del *tercer mundo*. Esta distinción se basa en el criterio del desempeño en cuanto a la producción. Denominaremos primer mundo a los países que tienen, en relación a los del tercer mundo, un mayor nivel del producto por persona, es decir, será el grupo de países relativamente ricos.

Los países que funcionaban con las reglas y organizaciones del socialismo eran denominados el «segundo mundo». A los ex países socialistas de Europa ahora se les llama «países en transición» (se supone hacia el capitalismo). Como nuestro objeto de estudio serán los países capitalistas de larga trayectoria, excluiremos de este objeto de estudio a los ex países socialistas de Europa del este y Rusia, así como también excluiremos a los actuales países socialistas, como la República Popular China, Corea del Norte, Cuba y Vietnam.

Según los datos del Banco Mundial (2001: cuadro 1), los primeros veinte lugares de los países capitalistas más ricos del mundo lo ocupan Alemania, Austria, Bélgica, Dinamarca, España, Finlandia, Francia, Italia, Holanda, Noruega, Portugal, Reino Unido, Suecia y Suiza, en Europa occidental; Estados Unidos y Canadá, en América del Norte; Japón en Asia; y, Australia y Nueva Zelanda en Oceanía. Este grupo formará el primer mundo en nuestra definición.

El tercer mundo incluye al resto de países del mundo capitalista. Más adelante se hará una separación del tercer mundo en dos grupos, según el criterio de su historia colonial, aquellos con un legado colonial importante y aquellos sin ese legado. Se mostrará que el segundo grupo es muy pequeño dentro del tercer mundo. Por lo tanto, en este capítulo se mantendrá el tercer mundo como una unidad para simplificar las comparaciones.

Las dos características que distinguen el primer mundo del tercer mundo, a lo largo del libro, serán las dotaciones factoriales y la desigualdad en la distribución de activos económicos y políticos. La primera se refiere a la cantidad de capital por trabajador. No existen mediciones estadísticas sobre esta relación, pero sería muy difícil disputar que esta es mayor en el primer mundo en comparación a la del tercer mundo.

En la segunda característica, la desigualdad en las dotaciones individuales de activos económicos se refiere a la concentración en tierras agrícolas, capital físico y capital humano. Tampoco en este caso existen mediciones estadísticas, pero sería muy difícil cuestionar que la concentración de estos activos es mayor en el tercer mundo que en el primer mundo. En el tercer mundo los latifundios son más frecuentes; la propiedad del capital físico está menos extendida, debido al poco desarrollo de los mercados de acciones empresariales; y la desigualdad en años de educación de la población adulta es también mucho mayor, si se le mide, por ejemplo, por la tasa de analfabetismo. En cuanto a los activos políticos, es evidente que los derechos ciudadanos están menos desarrollados en el tercer mundo, en comparación al primer mundo.

En cuanto a las regularidades sobre la producción y la distribución, algunas son propias del primer mundo, otras del tercer mundo y otras de ambos grupos. Las siete regularidades empíricas que se han podido establecer a partir de la literatura empírica internacional se muestran a continuación.

Regularidades por grupos de países: primer mundo versus tercer mundo

El primer grupo de regularidades se refiere a aquellas que son diferenciadas, unas son propias del primer mundo y otras del tercer mundo. Ellas son:

1. Existencia y persistencia del desempleo en el primer mundo

Los países del primer mundo operan con desempleo. Los desempleados son aquellos trabajadores que buscan empleo en las firmas y no lo consiguen. El desempleo es, ciertamente, una situación no deseada por los trabajadores. Las tasas de desempleo pueden subir o bajar pero son típicamente positivas. Es muy inusual que una economía capitalista opere sin desempleo. Las estadísticas internacionales así lo muestran. En el conjunto de países de Europea occidental, en el periodo 1960-2005, la tasa promedio anual ha oscilado entre 2% y 10%; en Estados Unidos estas tasas han variado entre 3% y 10% (Blanchard 2004: cuadro 1.5).

Es importante señalar una experiencia que han enfrentado los ex países socialistas desde que empezaron su transición al capitalismo. Entre las primeras cosas que aparecieron, junto con el desarrollo de los mercados, ha sido el desempleo, fenómeno que era casi desconocido en la sociedad socialista. Uno puede entender ahora por qué el historiador John Garraty llamó al desempleo la enfermedad del capitalismo.

2. Existencia y persistencia del desempleo y subempleo en el tercer mundo

En el tercer mundo se puede distinguir tres grupos de trabajadores. El primero es el que se encuentra empleado en las firmas a cambio de salarios; el segundo es el que busca ese tipo de empleo pero no lo consigue, que son los desempleados. Y existe un tercer grupo constituido por el grupo de trabajadores que crea su propio empleo. Ellos son los que se autoemplean en pequeñas unidades productivas, que es donde generan sus ingresos.

La característica del autoempleo es que el ingreso generado allí es, en promedio, inferior al salario que rige en el mercado laboral, para una calificación laboral similar. Al igual que el desempleo, el autoempleo no es tampoco una situación deseada por los trabajadores; es, más bien, una situación de segunda opción, siendo la primera obtener el empleo asalariado. Denominaremos, por eso, *subempleo* al caso del grupo de trabajadores que se encuentra en la situación de autoempleo y obtiene ingresos inferiores al salario del mercado para una misma calificación.

El desempleo es también característico del tercer mundo. La regularidad estadística es que la tasa de desempleo puede mostrar variaciones por países y periodos, pero usualmente es positiva. Los rangos de las tasas de desempleo no son muy distintas a las del primer mundo.

El autoempleo es, en cambio, típico del tercer mundo. La tasa de autoempleo entre los trabajadores de las ciudades es mucho mayor en el tercer mundo que en el primer mundo. Según los datos de la Organización Internacional del Trabajo (OIT), para el periodo 1980-2000, esta tasa fue, en promedio, de 12% para el primer mundo y de 40% para el tercer mundo, para una muestra de 17 y 58 países, respectivamente (OIT 2002: anexo 2, 62-64). Si se hiciera un recálculo tomando en cuenta también a la población rural, estas diferencias serían aun mayores, pues la tasa de autoempleo es mayor en el campo que en la ciudad en todas partes, pero en el tercer mundo la diferencia es claramente mayor que en el primero. No todo el autoempleo es igual al

subempleo; sin embargo, dado que no existen mediciones directas del subempleo, los tomaremos aquí como equivalentes.

En suma, la diferencia en la situación laboral entre el primer mundo y el tercer mundo es la enorme tasa de subempleo que existe en este último. Las diferencias no están tanto en las tasas de desempleo.

3. Existencia y persistencia de diferencias en ingresos entre grupos étnicos en el tercer mundo

En el tercer mundo, la mayoría de los países son multiétnicos. Las diferencias en el ingreso medio entre grupos étnicos son marcadas y persistentes. Estas diferencias son también jerárquicas, pues el orden de los grupos se mantiene en el tiempo. No es que un grupo étnico pase de ser pobre a rico, o viceversa, de un periodo a otro. Los grupos étnicos permanecen de manera sistemática o como grupos pobres o como grupos ricos. En los países que han tenido un legado colonial, los grupos étnicos que son descendientes de poblaciones aborígenes o esclavas son las que se ubican, en su mayoría, dentro del grupo pobre del país.

Los estudios empíricos que han calculado las desigualdades de ingresos entre grupos étnicos han encontrado que, en efecto, existen esas diferencias y son estadísticamente significativas. Estos estudios se refieren a varios países de América Latina, África y Asia para periodos variados de las últimas décadas (Darity y Nembhard 2000; Hall y Patrinos 2005; Psacharopoulos y Patrinos 1994; Silva 2001; y, Stewart 2001). En suma, la literatura internacional muestra que en la desigualdad de ingresos que se observa en el tercer mundo, la etnicidad de los grupos sociales cuenta.

Regularidades en el capitalismo mundial: primer mundo y tercer mundo

El segundo grupo de regularidades se refiere a aquellas que son válidas para todos los países capitalistas. Ellas son:

4. En el corto plazo (dada la capacidad productiva), las variables nominales y reales se mueven conjuntamente

Cuando los gobiernos mueven las variables nominales que controlan, como la cantidad de dinero, la tasa de cambio, la tasa de interés o el precio de un bien que produce una empresa pública, ¿se ven afectadas las variables reales de la economía, tales como el nivel de producción, nivel de empleo y salario real?

Los estudios empíricos tienden a dar una respuesta afirmativa a esta pregunta. Para los Estados Unidos, durante el periodo 1959-1996, y sobre datos trimestrales, un estudio encontró que las variaciones en la cantidad de dinero estaban correlacionadas positivamente —se movían en la misma dirección— con el nivel del producto, aunque el grado de correlación era moderado (Barro 1997: capítulo 18). Para el periodo 1970-1997, otro estudio también mostró correlaciones positivas entre esas variables nominales y reales en una muestra de diez países del primer mundo y quince del tercer mundo (Rand y Tarp 2002).

5. En el largo plazo (cuando aumenta la capacidad productiva), el nivel del producto y los salarios reales se mueven en la misma dirección

Cuando el nivel de producción de una economía capitalista crece, el salario real no disminuye ni se queda estancado, sino que también aumenta. El crecimiento económico de los países capitalistas, en la mayoría de los casos, no pauperiza a sus trabajadores asalariados.

En los Estados Unidos, por ejemplo, durante el periodo 1959-1973, el producto por trabajador creció a la tasa anual de 1.9% y el salario

real lo hizo casi a esa misma tasa (Barro 1997: capítulo 6). Para el caso de América Latina, la evidencia empírica para el periodo 1990-1999 muestra que de los once países que experimentaron un crecimiento sostenido en su nivel del producto total, en ocho de ellos, los salarios reales en la industria también crecieron de manera sostenida (OIT: 2000, cuadro 9a).

Regularidades sobre desarrollo económico entre los dos grupos de países

El tercer grupo de regularidades se refiere a las tendencias de largo plazo entre los dos grupos de países. A continuación les presentamos dichas tendencias.

6. Persistencia de las diferencias en el nivel de ingreso entre el primer mundo y el tercer mundo

El primer mundo, por la definición utilizada en este estudio, se compone de los veinte países capitalistas más ricos. Según los datos del Banco Mundial (2002: cuadro 1), los países del primer mundo tenían en 1999 un ingreso por persona de 26.000 dólares anuales, mientras que los del tercer mundo llegaban apenas a los 1.200, una diferencia de veinte veces. América Latina alcanzaba un valor de 3.800 dólares, que la hace la región más rica dentro del tercer mundo.

Usualmente estas cifras se someten a correcciones por las diferencias de precios que existen entre el primer y el tercer mundo, a fin de medir el ingreso en dólares de igual poder adquisitivo. Con esta corrección las diferencias cuantitativas se reducen, pero el orden de los países no cambia mucho. Se debe señalar, sin embargo, que los métodos de corrección hacen varios supuestos implícitos, como por ejemplo que en ambos mundos se consumen los mismos bienes —aunque en diferentes cantidades—, cuando tal vez lo más notable entre ambos mundos sea

precisamente que los bienes que se consumen son distintos. Aunque la medición mostrada anteriormente es imperfecta, debemos aceptarla.

Debido a que el primer mundo es, por definición, el grupo con mayores niveles de ingreso, la pregunta significativa es si esa diferencia ha persistido en el tiempo. La literatura internacional muestra que en las últimas cuatro a cinco décadas, tanto los países del primer mundo como los del tercer mundo han mostrado crecimiento en el nivel del producto. Pero las tasas de crecimiento en el tercer mundo no han sido mayores que las del primer mundo. Así, un estudio empírico, que ya es un clásico, encontró este resultado, utilizando una muestra de 114 países del primer y del tercer mundo para el periodo 1960-1990 (Barro y Sala-i-Martin 1995).

La evidencia empírica es que, en general, los países del tercer mundo no han podido alcanzar el nivel de ingreso que tienen los países del primer mundo. Según el historiador Angus Maddison (1995), desde 1850 no hay nuevos miembros en el club del primer mundo, siendo la única excepción Japón, que hoy se encuentra entre los países del primer mundo.

7. Existencia y persistencia en el mayor grado de desigualdad en los países del tercer mundo con respecto a los del primer mundo

El grado de desigualdad o más propiamente el grado de concentración del ingreso nacional en los grupos ricos de una sociedad se mide usualmente por el coeficiente de Gini, cuyo valor va desde cero —igualdad perfecta, todos los individuos tienen el mismo ingreso— hasta uno —concentración perfecta, un individuo tienen todo el ingreso nacional—. Empíricamente, sin embargo, es poco frecuente encontrar un coeficiente de Gini superior a 0,60 o por debajo de 0,20.

El grado de desigualdad en la distribución del ingreso nacional, medido por el coeficiente de Gini, mostró un valor promedio de 0,33 en los países el primer mundo y de 0,45 en los del tercer mundo durante el periodo 1950-1995 (Deninger y Squire 1996: cuadro 1). Estos datos

muestran que los países del primer mundo tienden a ser sociedades más igualitarias en comparación a las del tercer mundo. Sobre la desigualdad, esta es la primera regularidad.

Por otro lado, varios estudios han mostrado que los grados de desigualdad de los países tienden, en general, a cambiar muy poco a través del tiempo (Deninger y Squire 1996: cuadro 5; Li, Squire y Zou 1998). En consecuencia, se puede concluir que el orden en los grados de desigualdad entre los dos grupos de países tiende a mantenerse a través del tiempo. Sobre la desigualdad, esta es la segunda regularidad.

Tareas del libro

Cualquier teoría económica que intente explicar el funcionamiento del capitalismo tiene que dar cuenta de estas siete regularidades. Estas regularidades juegan el papel de evidencia empírica (el conjunto de datos denominados con la letra b en el capítulo 1) para refutar las teorías.

En el resto del libro se presentarán varias teorías que intentan explicar el capitalismo. Las predicciones de dichas teorías serán luego contrastadas contra estas regularidades y serán, así, sometidas al proceso de falsación.

De acuerdo a los principios epistemológicos establecidos en el primer capítulo, una teoría es sometida a la falsación a través de sus modelos. Una teoría falla cuando todos sus modelos fallan; un modelo falla cuando es refutada por algunas de las regularidades; por lo tanto, una teoría falla cuando todos sus modelos fallan en alguna de estas regularidades. Si una teoría no es refutada por ninguna de las regularidades no significa que ella sea verdadera, pero la teoría será aceptada provisionalmente, pues no existe razón alguna para rechazarla en esta etapa de la investigación. Este es el principio de la razón insuficiente aplicado a la falsación de las teorías. Utilizaremos estos principios a lo largo del libro.

Capítulo 3

La economía estándar

En la ciencia económica actual coexisten varias escuelas. Las principales son tres: neoclásica, clásica y keynesiana. Ciertamente, cada escuela tiene una teoría diferente sobre el funcionamiento del capitalismo. Ellas conforman lo que se puede denominar la *economía estándar*.

Según Thomas Kuhn, conocido físico y epistemólogo, las escuelas compiten al interior de una ciencia y como resultado de esa competencia queda solo una, que es la dominante o el paradigma. Esta escuela dominante se constituye en la «ciencia normal», pues establece no solo la explicación sobre la realidad social, sino también el campo y el método de análisis dominantes.

Aunque coexisten varias escuelas en la ciencia económica actual, se puede decir que la escuela neoclásica es la dominante; es el paradigma actual. El paradigma se reconoce fácilmente porque es el material que se utiliza en los centros educativos. La economía neoclásica es la que se encuentra en los textos universitarios de economía que se enseñan en el planeta. Es curioso, pero la economía se enseña en todas las universidades del mundo utilizando los mismos textos, al igual que la enseñanza de la física.

En este capítulo se presentan las proposiciones alfa de cada escuela; luego, se incorporan supuestos auxiliares para construir modelos de cada teoría; y, finalmente, se confrontan sus proposiciones beta con las regularidades empíricas presentadas en el capítulo anterior.

Teoría de la firma

Las tres escuelas utilizan una misma teoría sobre el comportamiento de las firmas. El supuesto primario es que la firma capitalista busca maximizar las ganancias. Para que esta teoría pueda ser operacional y refutable, introduciremos supuestos auxiliares para generar un modelo. Supondremos que la firma opera en mercados de competencia perfecta, esto es, donde ninguna firma tiene poder para fijar precios. Los precios son exógenos a la firma. Analizaremos el comportamiento de la firma en el mercado laboral.

Los capitalistas buscan hacer ganancias contratando trabajadores; es decir, la contratación de trabajadores no es el fin de la actividad de la firma sino un medio, el medio para generar ganancias. Para entender que la firma contrata trabajadores guiada por la motivación de la ganancia máxima, considere el caso de una firma típica. Considere que esta firma opera con ochenta trabajadores. ¿Por qué contrata ochenta y no otra cantidad? La teoría dice que, dado el salario del mercado y la dotación de capital de la firma, con ochenta trabajadores las ganancias serán mayores en comparación a si se emplea cualquier otra cantidad.

Pero, nuevamente, ¿por qué ochenta trabajadores llevan a la ganancia máxima? El salario es para la firma parte de sus costos. La ganancia de contratar trabajadores resultará de la cantidad total que se producirá con esa cantidad de trabajadores y, la venta de esa producción en el mercado del bien, al que llamaremos el bien B, al precio que rige en el mercado, menos los costos laborales totales. Esta diferencia se hace máxima con ochenta trabajadores.

¿Por qué con ochenta trabajadores? Necesitamos entender el proceso productivo. Se supone que a mayor cantidad de trabajadores, la cantidad producida del bien será mayor; pero la relación no será lineal, es decir, el doble de trabajadores no producirá el doble de producto, sino menos del doble. Este es el principio de los *rendimientos decrecientes factoriales* —de un factor que es variable— en el proceso productivo. El supuesto de que un factor de producción utilizado en

el proceso productivo en cantidades crecientes llegará, finalmente, a mostrar rendimientos decrecientes es considerado casi una verdad y por eso se le denomina principio, en lugar de supuesto. Si no fuera por los rendimientos decrecientes factoriales se podría producir los alimentos del mundo en una maceta de tierra.

El principio de los rendimientos decrecientes implica que el producto por trabajador o *productividad promedio* disminuirá cuando la cantidad de trabajadores aumenta. El rendimiento decreciente se expresa de manera intuitiva en ese dicho popular: «a más regimiento menos rendimiento». Si usted, amigo lector, puede construir un gráfico le ayudará a entender este principio. En el eje vertical mida la cantidad producida en un día en una firma y en el eje horizontal la cantidad de trabajadores empleada. Dibuje la curva del producto total, una curva creciente pero que cada vez crece menos. ¿Qué le dice esta curva con respecto al producto por trabajador?

El principio de los rendimientos decrecientes también implica que la contribución de un trabajador *adicional* a la cantidad producida, llamada *productividad marginal*, también disminuirá a medida que aumenta la cantidad de trabajadores. Si la firma contrata el trabajador 79, el producto aumentará en, digamos, 40 unidades diarias; si contrata el trabajador 80, el producto aumentará, pero en una cantidad menor, digamos, en 30 unidades, y con el trabajador 81 aumentará en, digamos, 20 unidades. Amigo lector: ¿puede usted incluir estos datos en su gráfico?

Una consecuencia importante a destacar del principio de los rendimientos decrecientes es que el producto marginal es menor que el producto promedio, pues el promedio puede bajar solo si el marginal es inferior. Un ejemplo le puede ayudar a entender esta relación: si la edad promedio en un grupo de personas es treinta años, para que el promedio disminuya, la persona que se una al grupo deberá tener una edad inferior a treinta años.

Hasta ahora hemos hablado de la productividad marginal *física* del trabajo. Esta cantidad se puede transformar en unidades monetarias

utilizando el precio del bien producido, que en este modelo es fijo para la firma. Así construimos el concepto de valor de la productividad marginal laboral, que es igual a la productividad marginal física laboral multiplicada por el precio del bien producido.

Ahora sí estamos listos para responder la pregunta sobre la lógica de la cantidad de trabajadores empleada por la firma. Si la firma contratara trabajadores hasta que el valor de la productividad promedio fuese igual al salario del mercado, ¿cuánto sería la ganancia? Cero, pues todo el producto se iría en pagar salarios. Las firmas no pueden seguir este criterio. En cambio, si la firma contrata trabajadores hasta que el valor de la productividad marginal laboral sea igual al salario de mercado, la ganancia será máxima.

Para poner un ejemplo, considere los datos de productividad física utilizados anteriormente; en adición, considere que el precio del bien B es igual a un sol por kilo, de modo que la productividad marginal física laboral será igual al valor de la productividad marginal laboral, y sea el salario de mercado igual a treinta soles diarios. El trabajador ochenta contribuye al aumento del ingreso total de la firma en treinta soles, que es igual a su contribución al aumento del costo total laboral, también de treinta soles. Como el trabajador 81 contribuirá menos al ingreso total —con veinte soles— que al costo laboral total, si lo contratara, la firma perdería esa diferencia; en cambio, el trabajador 79 contribuirá más a los ingresos totales —cuarenta soles— que al costo laboral total, y entonces si no lo contratara, la firma dejaría de ganar esa diferencia. Lo mismo sucede con los demás trabajadores, hasta llegar al trabajador ochenta.

La firma paga un salario uniforme, el salario del mercado, a todos los trabajadores. La firma contrata trabajadores hasta que el valor de la productividad marginal sea igual al salario del mercado. Esta igualdad ocurre en el trabajador ochenta y por eso contrata esta cantidad. Para los 79 trabajadores restantes, la productividad marginal está por encima del salario, y es así como la firma genera la ganancia máxima. Así podemos entender ahora por qué la firma contrata ochenta trabajadores. En suma, si la firma desea maximizar la ganancia total deberá contratar

ochenta trabajadores, que es la cantidad que iguala el valor de la productividad marginal laboral al salario nominal del mercado.

¿Qué pasaría con la ganancia si la firma pudiera pagar a cada trabajador el valor de la productividad marginal? La ganancia sería cero. En este modelo, la firma paga un mismo salario a todos los trabajadores, el salario del mercado. Si la firma pagara por encima del salario del mercado, estaría gastando en salarios más de lo necesario y no tendría ningún beneficio a cambio; si pagara por debajo, no podría conseguir trabajadores. También debe quedar en claro que si la firma iguala la productividad marginal laboral al salario de mercado, la productividad promedio del trabajo será necesariamente superior al salario.

¿Es así como actúan las firmas? Hace años un estudio mostró que los empresarios ingleses no conocían el término productividad marginal, entonces, no podrían igualarla a los salarios, y el estudio saltó a la conclusión de que la teoría era falsa. Esta no es forma de refutar empíricamente una teoría de la firma, la cual construye una firma abstracta. Serán las proposiciones beta de este modelo de la teoría las que nos dirán si la teoría es buena o no, si la firma abstracta se parece a la firma del mundo real, cosa que se mostrará más adelante. Por ahora es suficiente decir que la firma del mundo real no necesita utilizar estos conceptos ni hacer estos cálculos. Pero cualquier otro mecanismo de cálculo económico que utilice la firma real —de manera intuitiva, por ejemplo—, será equivalente al que acabamos de describir, si es que busca maximizar sus ganancias; es *como si* la firma del mundo real igualara el salario de mercado a la productividad marginal del trabajo.

La teoría dice que la firma busca maximizar ganancias. La teoría se refiere a la firma abstracta y concluye que si busca maximizar ganancias debe igualar la productividad marginal del trabajo al salario del mercado; pero la teoría no dice que la firma del mundo real hace esto. La prueba empírica consiste en saber si la firma abstracta se parece a la firma real, y no importa si la firma real sepa o no sepa cómo funciona la firma abstracta.

Este es el mismo problema del jugador de fútbol que busca hacer un gol desde una posición de tiro libre. Si busca el gol, el jugador pateará la pelota buscando maximizar la curvatura de su trayectoria. Este problema es matemático, pero el jugador no resuelve las ecuaciones antes de patear, sino resuelve el problema de manera intuitiva —con su talento—. Pero lo que haga es equivalente a resolver las ecuaciones. Los grandes futbolistas no necesitan tener un doctorado en Matemáticas, y en efecto no lo tienen. Pero el teórico necesita utilizar las matemáticas para explicar el comportamiento del futbolista y poder predecir cómo pateará el balón a diferentes distancias del arco.

Otra precisión se hace necesaria aquí. Habrá notado usted, amigo lector, que el término que se utiliza es productividad del *trabajo* y no del *trabajador*. La razón es que el trabajador no tiene una productividad personal. La cantidad de producto que un trabajador individual contribuye a producir dependerá de la cantidad de trabajadores y la cantidad de máquinas con las que labora en la firma. La productividad es resultado de todas las interacciones en el proceso productivo. La productividad no es individual, es social. No sería posible entonces escribir en el documento de identificación del trabajador, junto a sus datos personales, el dato sobre el valor de su productividad. Se puede escribir su profesión, pero eso no es su productividad.

El salario se puede medir de dos formas, como salario nominal o salario real. El salario nominal es el precio del trabajo medido en unidades monetarias —treinta soles por día—. El salario real es el precio del trabajo medido en unidades físicas de un bien; es el poder de compra del salario nominal, y se le mide dividiendo el salario nominal entre el precio del bien. Si el salario nominal es de treinta soles por día y el precio del bien, digamos arroz, fuese de dos soles por kilo, el salario real sería igual a quince kilos de arroz.

Utilizaremos el concepto de salario real en el análisis del mercado laboral. La condición de equilibrio en la contratación de trabajadores por la firma individual es entonces la igualación de la productividad marginal física laboral con el salario real. De esta condición de equilibrio,

se puede derivar una relación empírica que es observable y que puede refutar el modelo: si el salario real baja, la firma tendrá el incentivo de contratar una mayor cantidad de trabajadores; si no lo hace, el salario real será menor que la productividad marginal del trabajo y la firma no estaría maximizando sus ganancias.

A la relación inversa entre salarios reales y cantidad empleada se le denomina la curva de demanda de trabajo de la firma. La curva de demanda del mercado laboral será simplemente la suma de las cantidades de trabajadores que cada firma desea contratar para cada salario real; por lo tanto, también será una curva decreciente. La posición de la curva está determinada por la cantidad de capital del conjunto de firmas. Si el stock de capital aumenta, aumentará la productividad promedio laboral y, por lo tanto, también la productividad marginal. La curva de demanda de trabajo estará ahora en una posición más alta, pues al mismo salario real las firmas desearán contratar más trabajadores. Esta característica de la demanda de trabajo será la misma en las tres teorías que se presentan a continuación.

La sociedad neoclásica

La sociedad neoclásica se origina en los trabajos de Adam Smith (1776) y Leon Walras (1883). El mundo abstracto neoclásico se construye haciendo supuestos específicos sobre los componentes principales de un sistema económico, tal como estos fueron definidos en el primer capítulo.

Proposiciones alfa

En esta sociedad, los individuos intercambian bienes a través del mercado. El intercambio de mercado se define por dos características: es voluntario y los que participan tienen que estar dotados de bienes en propiedad privada. El que no participa en un mercado particular es porque no desea hacerlo o porque, deseando hacerlo, no tiene suficientes

bienes para hacer el intercambio. Como los individuos actúan guiados por la motivación del interés personal, en el intercambio de mercado los individuos actúan de manera egoísta, cada uno busca obtener la mayor ventaja personal del intercambio.

Los precios nominales de los bienes —cantidad de dinero por unidad del bien— son flexibles y se pueden mover libremente en el mercado; por lo tanto, si al precio del mercado no todos pueden intercambiar las cantidades que desean —sea comprar o vender a ese precio—, el precio se moverá, y se moverá hasta que todos puedan intercambiar las cantidades deseadas. A este tipo de mercado se le llama *mercado walrasiano*, y al precio que rige en este tipo de mercado se le llama *precio walrasiano*. El precio walrasiano es el que «limpia» el mercado; es decir, a este precio nadie se queda sin poder transar las cantidades que desea.

Estas características de la sociedad neoclásica se pueden expresar en las proposiciones alfa siguientes:

Sobre el contexto institucional. Las reglas de funcionamiento de la sociedad incluyen el respeto a la propiedad privada de los activos económicos con los cuales están dotados los individuos; el intercambio de bienes opera a través del mercado, y los mercados son walrasianos. Las organizaciones que existen son los hogares, las firmas y el gobierno. El gobierno ofrece el dinero.

Sobre la racionalidad de los agentes. Los individuos actúan guiados por la motivación del interés personal.

Sobre las condiciones iniciales. La condición inicial de la sociedad es que la dotación inicial de factores —stock de capital por trabajador— es tal que la productividad laboral es suficientemente alta como para que los salarios del mercado se guíen por ella.

¿Será que el mundo real funciona como este mundo abstracto? Necesitamos construir modelos de la sociedad neoclásica para responder esta pregunta. Considere el siguiente modelo. Supondremos, en primer lugar, dos tipos de capital que ingresan al proceso productivo:

el capital físico —máquinas y herramientas— y el capital humano, que se refiere al stock de conocimientos productivos que se encuentra incorporado en los trabajadores. En segundo lugar, existen dos clases sociales: capitalistas y trabajadores. Los capitalistas están dotados de capital físico con el cual operan sus firmas y también están dotados de capital humano —mano de obra calificada—. Los trabajadores están dotados de capital humano —mano de obra calificada— y también de cantidades de dinero como stock. El stock de dinero de la economía es igual al emitido por el gobierno.

En tercer lugar, existen tres mercados: el mercado del bien B —el único bien que se produce en la economía—, del trabajo y del dinero. Estos mercados operan bajo la forma de competencia perfecta, lo cual significa que nadie tiene poder para fijar el precio. En cada mercado, el precio walrasiano se determina por la interacción de todos los participantes, compradores y vendedores.

Finalmente, supondremos que la economía es cerrada y estática. El primer supuesto dice que se ignorará el efecto del intercambio con otras economías; y el segundo dice que las soluciones del mercado sobre los precios y las cantidades transadas —el valor de las variables endógenas— se repetirán periodo tras periodo, mientras las variables exógenas se mantengan fijas.

Mercado laboral

Veamos ahora cómo funciona cada mercado. Comencemos con el mercado laboral. Debido a que los trabajadores son homogéneos —todos tienen la misma calificación—, habrá un solo mercado laboral, con un único salario. Existe una cantidad dada de trabajadores que desea intercambiar su trabajo por el bien B y están dispuestos a trabajar a cualquier salario. Sobre la demanda, necesitamos entender, primero, el comportamiento de la firma individual para luego comprender el comportamiento agregado de las firmas en el mercado laboral.

Con el estudio del mercado laboral buscamos entender la determinación de dos variables endógenas: el precio y la cantidad transada; es decir, el salario y el nivel del empleo. La demanda de trabajo se expresa, como vimos anteriormente, en una curva de demanda que es decreciente en relación a los salarios reales. Se supondrá que la oferta es una cantidad fija de trabajadores, quienes están dispuestos a trabajar a cualquier salario.

Allí donde se cruzan la curva de demanda de trabajadores con la curva de oferta se determinará el salario real del mercado. El precio de solución del mercado laboral será el salario real walrasiano, pues a este precio ningún comprador o vendedor se quedará sin poder realizar el intercambio deseado. El mercado laboral opera como un mercado walrasiano. Una vez establecido dicho precio, nadie tiene el poder ni el deseo de modificar este precio. La condición de equilibrio implica desempleo igual a cero.

El funcionamiento del mercado laboral se puede resumir en un gráfico muy simple. Considere la figura 3.1, donde en el eje vertical se mide el salario real (w) y en el eje horizontal la cantidad de trabajadores (Q_h). La curva decreciente representa la demanda y la línea vertical la oferta. El mercado laboral es walrasiano y entonces el salario de equilibrio es el que corresponde al cruce de las dos curvas, en el punto E. A este precio nadie se queda sin tranzar las cantidades que desea intercambiar; en particular, la situación de equilibrio, el punto E, implica que todos los trabajadores están empleados.

El área marcada con W^o en la figura 3.1 indica la masa de salarios reales que va a los trabajadores y el área marcada con P^o, el total de ganancias de las firmas. La suma de las dos áreas (W^o+P^o) tiene que ser igual al producto total, pues el producto total solo se puede distribuir entre estas dos clases; luego, en el gráfico, el producto total es igual al área debajo del segmento HE.[4]

[4] La propiedad de que el área debajo de la curva de la productividad marginal HE es igual al producto total se puede ilustrar con un ejemplo. Si el empleo de un trabajador

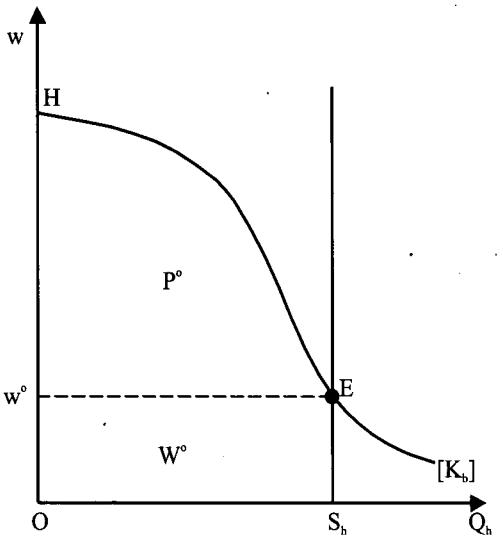

Figura 3.1 Producción y distribución en una sociedad neoclásica

¿Cómo es posible que existan ganancias si las firmas operan en un mercado de competencia perfecta?, ¿por qué esas ganancias no son destruidas por la entrada de otras firmas? En el corto plazo, dado el stock de capital de la economía y dada su distribución entre la clase capitalista, las ganancias no pueden ser destruidas por la competencia.

En el largo plazo, cuando el stock de capital aumenta, las ganancias podrían ser destruidas. Pero se introducirá ahora un nuevo supuesto que evitará tal resultado. Se supondrá que las firmas son heterogéneas. Se diferencian entre ellas no solo por las distintas dotaciones en su stock de capital, sino también por el talento empresarial de los capitalistas y por la localización de las firmas. Este último supuesto implica un

---

da lugar a una producción igual a cien unidades, es decir agrega cien unidades a la producción, y el de un segundo trabajador agrega a la producción sesenta unidades, y el de un tercer trabajador agrega cuarenta unidades (los rendimientos son decrecientes), queda claro que el producto total de emplear tres trabajador es es igual a doscientas unidades, es decir, es igual a la suma de las productividades marginales de los tres trabajadores.

acceso diferenciado de las firmas a los mercados, a los bienes públicos —como infraestructura— y a los recursos naturales, como agua y tierras de distinta calidad. Así, se supone que cada firma tiene su curva particular de productividad laboral y de la correspondiente curva de demanda laboral.

La curva de la demanda laboral del mercado es entonces la agregación de curvas heterogéneas de demanda individual. Esta característica se mantendrá también en el largo plazo. Con este supuesto adicional, el modelo predice que la competencia no podrá destruir las ganancias ni en el corto plazo ni en el largo plazo. Esta predicción empírica es consistente con la existencia y persistencia de las ganancias de las firmas del mundo real.

La figura 3.1 nos muestra así una representación gráfica del resultado del proceso productivo: la producción y la distribución de bienes entre las clases sociales en una sociedad neoclásica. El mercado laboral está en equilibrio, pues ningún actor social tiene el poder o el deseo de cambiar esta solución.

Equilibrio general

Si el mercado de trabajo está en equilibrio, ¿estará el mercado del bien B también en equilibrio? La respuesta es afirmativa. La cantidad demandada del bien B viene de los trabajadores y es entonces igual a la masa salarial real. La cantidad ofrecida al mercado por las firmas es igual a la cantidad producida menos la parte que los capitalistas retienen como ganancias, que también es igual a la masa salarial real. Luego, el mercado del bien B estará necesariamente en equilibrio como resultado de que el mercado laboral está en equilibrio.

Este resultado puede sorprenderlo, amigo lector. Pero es uno de los principios básicos del sistema de mercado. Se llama la *Ley de Walras*. Si existen dos mercados walrasianos en la economía, y uno de ellos está en equilibrio, el otro estará necesariamente en equilibrio. La razón reside en dos características del sistema de mercado: (a) la gente no puede

comprar algo sin vender algo; (b) el ingreso de un grupo es el gasto de otro grupo. Por lo tanto, gastos e ingresos tendrían que ser iguales en el agregado. Esto es lo que acabamos de demostrar en el caso de dos mercados, laboral y del bien B. Amigo lector: extienda el principio a tres mercados y luego generalice a «n» mercados, y así diviértase un poco con este principio básico.

Nos falta analizar el tercer mercado, el del dinero. ¿Estará también en equilibrio? No. Veamos por qué. Los trabajadores están dotados de cantidades de dinero, que los mantienen en su poder. ¿Qué hace ese dinero ocioso en su poder en lugar de estar ya en el estómago del trabajador como consumo del bien B? El supuesto que hacemos ahora es que los trabajadores tienen una demanda por dinero para cubrirse de todas las eventualidades que se les presenten en el mercado. Sus salarios los reciben, digamos, mensualmente pero sus gastos son diarios. Necesitan entonces conservar un stock de dinero, como un inventario. Es costoso para un individuo tener dinero ocioso en su poder, pues podría dedicarlo a comprar bienes de consumo. Pero el beneficio está en no perder una transacción que le es favorable. Por eso todos llevamos siempre una cantidad de dinero en la cartera o en el bolsillo, aunque tengamos tarjetas de crédito.

El dinero mismo también tiene dos valores, el nominal y el real. El nominal es el que se expresa en el mismo billete o moneda, en unidades de la propia moneda —un billete de cincuenta soles—. El valor real se refiere al poder de compra del dinero, a su valor en unidades del bien B. ¿Cuántos kilos de arroz se puede comprar con el billete de cincuenta soles? Dado el precio nominal del bien B —dos soles por kilo de arroz—, el poder de compra del billete de cincuenta soles queda determinado —veinticinco kilos de arroz—. El individuo que tiene en su billetera o cartera un billete de cincuenta soles como inventario, está andando con diez kilos de arroz, o con una botella de pisco, sin consumirlos. Si el precio del bien B se dobla, con el mismo billete se puede comprar solo la mitad de la cantidad de antes. El dinero real es entonces el dinero nominal dividido entre el precio del bien B.

¿De que factores dependerá la cantidad demandada de dinero? El supuesto de la teoría de la demanda de dinero es que esta depende del ingreso real: a mayor ingreso real, los trabajadores necesitarán mayores cantidades reales de dinero en su poder. Para el agregado de trabajadores, entonces, cuanto mayor es la masa real salarial, mayor será la cantidad de inventarios de dinero real que ellos necesitan. Por lo tanto, dada la masa salarial real, si el precio del bien B se dobla, los trabajadores necesitarán el doble de stock de dinero nominal para mantener el mismo stock real de dinero.

¿De dónde viene la oferta de dinero? En esta economía abstracta no tenemos bancos, de modo que el único dinero, es decir, el único medio de pago, es el dinero que emite el gobierno. ¿Qué factores determinan la cantidad de dinero de la economía? Supondremos que la cantidad de dinero en la economía está exógenamente determinada.

Luego, si el mercado laboral ha determinado la masa salarial real, la cantidad demandada de dinero real también ya está determinada. Debido a que la cantidad nominal de dinero en la economía está determinada exógenamente, el precio del bien B tendrá que cambiar para que la cantidad de dinero real sea igual a la cantidad demandada. El precio del bien B, que se denominará el nivel de precios de la economía, se determina endógenamente en el mercado de dinero.

La solución del equilibrio general es entonces secuencial. Primero se resuelven las variables reales de la producción y la distribución. Una vez conocidos estos valores, se determina el nivel de precios. Conocido el nivel de precios, habrá un salario nominal que sea consistente con el salario real ya determinado en el mercado laboral. El mercado laboral constituye el núcleo del sistema, pues una vez que el salario real queda determinado en este mercado, las demás variables endógenas se determinarán por implicancia.

Al nivel del producto de la sociedad se le llamará el ingreso nacional. Así, el ingreso nacional y su distribución entre las dos clases sociales que se representaron en la figura 3.1 se pueden representar ahora por las siguientes ecuaciones:

$$Y = P + W$$
$$= P + w S_h$$

Aquí Y representa el ingreso nacional; P es la parte del ingreso nacional que va a los capitalistas como ganancias; W es la masa salarial que va a los trabajadores como salarios; la masa salarial es igual al producto de multiplicar el salario real (w) por la cantidad de trabajadores (S_h). El desempleo es cero, pues el mercado laboral es walrasiano.

Falsación empírica

Debemos ahora confrontar las implicancias empíricas del modelo neoclásico que acabamos de presentar con las regularidades empíricas del capitalismo establecidas en el segundo capítulo. El modelo puede ser contrastado solo con las regularidades 1, 4 y 5.

Con respecto a la primera regularidad, el modelo falla. La condición de equilibrio de un mercado walrasiano predice que todo equilibrio será sin desempleo. Con respecto a la cuarta regularidad, el modelo también falla. El equilibrio general es secuencial, de modo que las variables reales se determinan antes de las variables nominales y con independencia de estas. Es decir, cambios en variables nominales no tienen efecto sobre las variables reales. Así, un aumento en la cantidad de dinero, solo aumentará el nivel de precios, pues no afectará ninguna variable real; es decir, el dinero es neutral.

Con respecto a la quinta regularidad, el modelo no falla. En el largo plazo, cuando el stock de capital aumenta, la productividad laboral aumenta, y la curva de demanda de trabajo se eleva (en la figura 3.1 la curva de demanda se desplaza hacia fuera) y el salario real aumenta. Luego, el producto total y el salario real aumentan.

El modelo neoclásico falla en explicar dos regularidades, la primera y la cuarta. La pregunta ahora es si con otros modelos de la teoría neoclásica se puede lograr explicar estas dos regularidades. La respuesta es negativa. Cualquier modelo neoclásico va a predecir pleno empleo

y neutralidad del dinero, pues estas son las predicciones empíricas que se derivan de las condiciones de equilibrio con mercados walrasianos. La familia de modelos de la teoría neoclásica falla en explicar esas dos regularidades empíricas del capitalismo.

En suma, la teoría neoclásica es refutada por algunos hechos de la realidad y no puede ser aceptada. La sociedad abstracta neoclásica no es una buena aproximación de la sociedad capitalista real.

La sociedad clásica

La sociedad clásica se origina en los trabajos de David Ricardo (1821) y Carlos Marx (1867). Los supuestos primarios de la teoría clásica, así como las diferencias con la teoría neoclásica, se refieren a las normas institucionales y a las condiciones iniciales de la sociedad.

Sobre las normas institucionales, la economía clásica supone que el mercado de trabajo funciona con un piso de salario real, exógenamente determinado, y que se llama el «salario de subsistencia». Con este ingreso los trabajadores pueden asegurar el consumo de una canasta de bienes considerada de subsistencia, que cubre necesidades fisiológicas y sociales.

Sobre las condiciones iniciales, la teoría supone, primero, que las dotaciones factoriales de la sociedad —cantidades de capital por trabajador— son tales que la productividad laboral es baja. En particular, la productividad marginal del trabajo en el nivel de pleno empleo es inferior al salario de subsistencia. Finalmente, la teoría supone desigualdad en la dotación inicial de activos económicos entre los individuos que implica la existencia de dos clases sociales: capitalistas y trabajadores.

El proceso de producción y distribución tiene sus particularidades en la sociedad clásica. Los capitalistas tienen el poder de establecer la duración de la jornada de trabajo —digamos cuarenta horas semanales—. El periodo de producción del bien B es, digamos, una hora. El nivel del producto semanal será entonces de cuarenta veces más. La

producción es lineal con respecto al tiempo: el doble del tiempo da el doble del producto.

Para un nivel de empleo dado en una firma, la productividad promedio semanal del trabajo es, digamos, cien unidades del bien B. Supongamos que cada trabajador necesita recibir como salario de subsistencia cincuenta unidades semanales. Esta cantidad se produce, entonces, en la mitad de la duración de la jornada, en veinte horas semanales. Este periodo se llama el *trabajo necesario*. En las restantes veinte horas, que se llama el *trabajo excedentario*, se producen cincuenta unidades que van a los capitalistas como ganancias. En este ejemplo, la *tasa de explotación* del trabajo, la relación trabajo excedentario sobre trabajo necesario, es igual a 100%. Es decir, por cada hora que el trabajador labora para su subsistencia, debe laborar otra hora para el capitalista.

Ciertamente, si no existe el trabajo excedentario, la firma no puede obtener ganancias. También es evidente que si la productividad promedio del trabajo es muy baja (cincuenta unidades en lugar de cien en el ejemplo anterior), el trabajo excedentario puede no existir; en este caso, las firmas no podrían obtener ganancias y el sistema capitalista no podría funcionar. Aquí supondremos que la productividad laboral es suficientemente alta como para generar ganancias.

Equilibrio general

Introduzcamos algunos supuestos auxiliares para construir un modelo clásico. Para facilitar la exposición, podemos introducir los supuestos auxiliares que utilizamos en el modelo neoclásico. Los capitalistas están dotados de capital físico y de capital humano y los trabajadores de capital humano y de tenencias de dinero. Existe un solo tipo de capital humano y entonces un único mercado laboral. Consideremos tres mercados, del bien B, laboral y del dinero. Los mercados son de competencia perfecta y la economía es cerrada y estática.

En el mercado laboral, supondremos que la duración de la jornada de trabajo semanal está exógenamente determinada. Dado el stock de

capital de la firma, la productividad semanal del trabajo depende de la cantidad de trabajadores. La firma contrata trabajadores guiada por la lógica de la ganancia máxima; es decir, contrata trabajadores hasta que la productividad marginal laboral sea igual al salario real. Luego, la demanda de trabajo de la firma será como en el modelo neoclásico: una curva decreciente que indica que a salarios reales menores, la cantidad demandada de trabajo será mayor.

La curva de demanda de trabajo del mercado se obtiene agregando las curvas individuales de las firmas. Dado el salario real de subsistencia, el mercado laboral solo determina la cantidad de empleo, la cual es determinada por la demanda. Esta cantidad demandada es inferior a la cantidad ofrecida de trabajadores; es decir, al salario de subsistencia, existe exceso de oferta laboral. Como el salario real está exógenamente determinado, no puede bajar para limpiar el mercado. El mercado laboral es no walrasiano. El equilibrio del mercado laboral será con desempleo. No existe ningún agente que tenga el poder y el deseo de cambiar esta situación.

El mercado laboral en este modelo clásico se representa en la figura 3.2. En el eje vertical se mide el salario real y en el horizontal la cantidad de trabajadores. La curva de demanda es decreciente y la de oferta es una línea vertical. Pero ahora el salario real está dado exógenamente, por el símbolo w*. A este salario real las firmas demandarán la cantidad D_h solamente, lo cual implica una cantidad de trabajadores desempleados. Esta es una situación de equilibrio, pues el salario real no puede bajar. La cantidad producida es igual al área debajo del segmento HE, el cual se distribuye entre las dos clases sociales, como salarios W y como ganancias P.

¿Dónde está el trabajo necesario y el excedentario en este equilibrio? Como la cantidad de empleo se determina por la condición de que la productividad marginal laboral debe ser igual al salario real de subsistencia, a ese nivel de empleo se determinará también el nivel de producto por trabajador. Y como la productividad marginal tiene que estar por debajo de la productividad media, esta diferencia nos dará

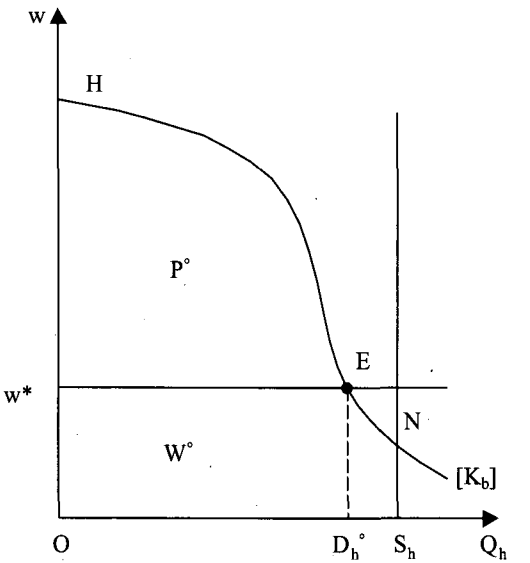

Figura 3.2 Producción y distribución en una sociedad clásica

indicaciones del trabajo necesario. Es decir, el trabajo necesario como proporción del trabajo total es igual a la productividad marginal como proporción de la productividad media, o también igual al salario real como proporción de la productividad media.

Se puede mostrar también que la relación de las ganancias sobre masa salarial (P/W) es igual a la relación trabajo excedentario sobre trabajo necesario, es decir, es igual a la tasa de explotación. Solo hay que notar que el producto semanal se puede descomponer en ganancia y masa salarial así: la ganancia es igual al producto horario multiplicado por el tiempo de trabajo excedentario y la masa salarial es igual al producto horario multiplicado por el tiempo de trabajo necesario. En suma, la igualdad entre la productividad marginal y el salario real distribuye el producto total entre trabajadores y capitalistas de una manera que es equivalente a la regla del trabajo necesario y excedentario.

En el mercado del bien B, la cantidad demandada es igual a la masa salarial y la cantidad ofrecida es igual al producto total menos las ganancias,

que es también igual a la masa salarial. Luego, si el mercado laboral está en equilibrio, el mercado del bien B también estará en equilibrio.

¿Qué sucede en el mercado de dinero? Suponemos que la teoría de la demanda de dinero es como en el modelo neoclásico desarrollado anteriormente. Luego, una vez conocida la masa salarial real, la cantidad demandada de inventarios reales queda también determinada. Dada la cantidad de dinero en la economía, el nivel de precios de equilibrio se determinará en el mercado de dinero. A este nivel de precios de equilibrio, el salario nominal se ajustará hasta que el salario real sea el de subsistencia.

En suma, el equilibrio general se compone de tres mercados, dos son walrasianos —del bien B y el monetario— y uno es no walrasiano —laboral—. Entre los dos mercados walrasianos uno es redundante, por la ley de Walras. Entonces se necesita resolver solo dos mercados, el laboral y, digamos, el de dinero. Pero el mercado de dinero no afecta al mercado laboral; entonces, la solución es secuencial, primero se resuelve el mercado laboral, para determinar el nivel de empleo, y luego el mercado monetario, para determinar el nivel de precios. El mercado laboral constituye el núcleo del modelo. Una vez que el nivel de empleo queda determinado en el mercado laboral, las otras variables endógenas se determinan por implicancia.

Como el salario de mercado sirve para cubrir el costo de subsistencia de los trabajadores, la masa salarial no puede ser considerada parte del producto neto de la sociedad. El producto de la sociedad es el excedente económico que resulta del trabajo excedentario.

En este modelo, el excedente económico es apropiado íntegramente por los capitalistas, como ganancias, debido a la propiedad que tienen del capital físico. Si la propiedad fuese de los trabajadores, ellos recibirían todo el producto, parte como salarios y parte como ganancias. En este caso, el ratio entre trabajo excedentario y trabajo necesario sería una medida de su autoexplotación. En este modelo simple, donde existe exceso de oferta laboral, el salario de mercado será siempre igual al de subsistencia.

Falsación empírica

Es hora de confrontar este modelo clásico con la realidad. Las predicciones empíricas de este modelo se pueden contrastar solo contra tres regularidades empíricas del capitalismo: 1, 4 y 5.

Con relación a la primera regularidad, el modelo falla. El modelo explica la existencia del desempleo, pero no puede explicar su persistencia. Un aumento sostenido en la demanda de trabajo, originada en un aumento sostenido en el stock de capital, llevaría a un nuevo equilibrio con pleno empleo. Una reducción importante en la oferta laboral también haría desaparecer el desempleo. En ambos casos, el mercado laboral se transformaría en un mercado walrasiano y, como consecuencia, el modelo clásico en un modelo neoclásico.

Con relación a la cuarta regularidad, el modelo falla. El equilibrio general es secuencial, donde las variables reales se determinan independientemente de las variables nominales. El dinero es neutral.

Con relación a la quinta regularidad, el modelo falla. En el largo plazo, al aumentar el stock de capital, aumentará la productividad laboral y con ello la demanda de trabajo; sin embargo, en el nuevo equilibrio, el salario real no aumentará debido al exceso de oferta laboral. El producto total aumenta, pero el salario real permanece fijo. Para aumentos significativos del stock de capital, que implicará aumentos igualmente significativos en la demanda de trabajo, el exceso de oferta laboral desaparecerá; en este caso, el modelo clásico llegará a ser un modelo neoclásico, y a partir de allí habrá consistencia con esta regularidad.

En suma, el modelo clásico falla en todos los casos. Cualquier otro modelo clásico llevaría a los mismos resultados, pues el equilibrio general será siempre con un salario real de subsistencia como piso en el mercado laboral. La familia de modelos de la teoría clásica tendrá las mismas predicciones empíricas. Luego, la teoría clásica es refutada por los datos de la realidad y no puede ser aceptada. La sociedad clásica no es una buena aproximación de la sociedad capitalista real.

La sociedad keynesiana

La sociedad keynesiana se origina el en trabajo de John Maynard Keynes (1936). Los supuestos básicos de la teoría keynesiana y la diferencia con la teoría neoclásica se refieren también a las normas institucionales. Una proposición alfa es que algunos precios nominales son exógenamente determinados. Entre estos precios está el salario nominal, el cual está determinado inicialmente por encima del salario walrasiano. Así, el mercado laboral opera inicialmente con exceso de oferta laboral.

Agreguemos ahora algunos supuestos auxiliares para construir un modelo keynesiano. Supondremos tres mercados, el del bien B, el laboral y el del dinero. Habrá dos clases sociales, capitalistas y trabajadores. Existe un único nivel de capital humano y por lo tanto un único mercado laboral. Los mercados son de competencia perfecta y la economía es cerrada y estática.

En su búsqueda de la ganancia máxima, y dado su stock de capital, la firma contratará trabajadores hasta que el valor de la productividad marginal del trabajo sea igual al salario nominal de mercado. El precio del bien B es también exógena para la firma individual. De este comportamiento de la firma se deriva una relación decreciente en la demanda de trabajadores: dado el precio del bien B, a mayores salarios nominales, la cantidad demandada será menor. Otra relación que se deriva es: manteniendo fijo el salario nominal, cuanto más alto es el precio del bien B, mayor será el valor de la productividad marginal, y por lo tanto mayor será la cantidad demandada de trabajadores.

Equilibrio general

La demanda de trabajo en el mercado laboral se obtiene por la simple agregación del comportamiento individual de las firmas. Dado el precio del bien B, el mercado laboral determinará el nivel de empleo por la intersección de la curva de demanda laboral con el salario nominal fijo. Ese nivel de empleo determinará la masa salarial real. Luego, se

conocerá en el mercado de dinero la cantidad demandada de dinero, y como el stock de dinero de la economía está dado, el nivel de precios quedará determinado. Pero, ya se había supuesto un nivel de precios para resolver el nivel de empleo, que puede ser distinto al que se ha obtenido ahora.

Bueno, lo único que estamos mostrando es que para determinar el nivel de empleo se necesita conocer el nivel de precios y para conocer el nivel de precios se necesita conocer el nivel de empleo; es decir, la solución de ambas variables endógenas es simultánea. Como se puede ver, las variables reales y nominales interactúan en este modelo. Como tenemos dos ecuaciones, una es la demanda de trabajo y la otra es la demanda de dinero, podemos esperar que tengamos una única solución de las dos variables endógenas.

Resuelto el nivel de empleo y el nivel de precios, los mercados monetario y laboral estarán ambos en equilibrio. Estos dos mercados constituyen el núcleo del modelo. Dado el salario nominal, con estos valores de equilibrio se determinará el salario real, la masa salarial real, el nivel del producto, las ganancias y el stock de dinero real. El mercado restante, el del bien B, también estará necesariamente en equilibrio. La cantidad demandada del bien B viene de los trabajadores y es entonces igual a la masa salarial real, mientras que cantidad ofrecida es igual al producto total menos las ganancias retenidas por los capitalistas, que es también idéntico a la masa salarial real.

En suma, existen tres mercados en este modelo keynesiano, donde el mercado laboral es no walrasiano y los otros dos son walrasianos. Luego, entre estos dos rige la Ley de Walras: si el mercado de dinero está en equilibrio, el del bien B también lo estará. Para llegar al equilibrio general es suficiente resolver dos mercados, uno walrasiano y otro no walrasiano, de manera simultánea. Aquí estos mercados son el mercado de dinero y el mercado laboral. Ambos mercados conforman el núcleo del equilibrio general, pues una vez resueltas las variables endógenas de estos dos mercados, el valor de las variables endógenas restantes del equilibrio general se obtienen por implicancias solamente.

Así hemos llegamos al equilibrio general. Podemos representar el equilibrio del mercado laboral en la figura 3.3. En el eje vertical se mide el salario real y en el horizontal la cantidad de trabajadores. La curva de demanda es decreciente y la curva de oferta es fija. El salario real —la relación salario nominal sobre nivel de precio— se determina endógenamente, pues el nivel de precios se determina por la interacción de variables reales y monetarias. Es decir, el salario real no se determina solo por las variables del mercado laboral. A este salario real existe exceso de oferta laboral, pero es un salario de equilibrio, pues no hay actor social que quiera y pueda cambiar este resultado. El mercado laboral es un mercado no walrasiano.

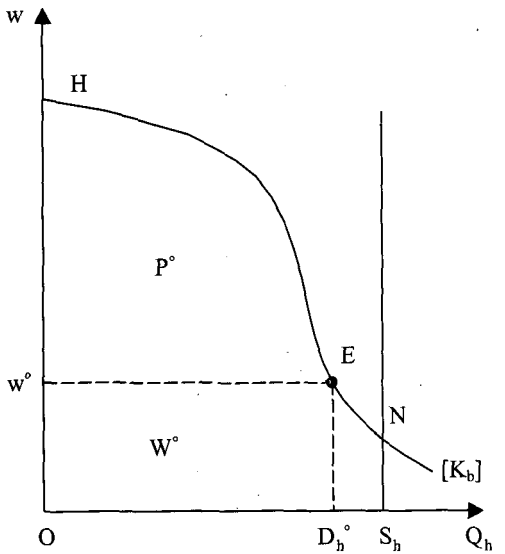

Figura 3.3 Producción y distribución en una sociedad keynesiana

El nivel del producto de la economía es el ingreso nacional Y. Su distribución entre salarios y ganancias se puede ver en la figura 3.3 por las áreas W y P. En forma de ecuaciones, la solución de producción, distribución y desempleo se puede representar así:

$$Y = P + w D_h$$
$$S_h = D_h + U$$

Debido a que el equilibrio general keynesiano es con desempleo, ahora existe una diferencia entre la cantidad de trabajadores S_h y la cantidad demandada D_h, que es igual a la cantidad desempleada U.

Falsación empírica

Las predicciones de este modelo keynesiano se pueden confrontar con las regularidades del capitalismo establecidas en el segundo capítulo. Esta confrontación se puede llevar a cabo solo con las siguientes tres regularidades: 1, 4 y 5.

Con respecto a la cuarta regularidad, el modelo no falla. En el equilibrio general, las variables reales y monetarias interactúan, esto es, el nivel de precios y el nivel de empleo se determinan simultáneamente. El dinero es exógeno. Luego, un aumento en la cantidad de dinero llevará al sistema a un nuevo equilibrio general, con un nuevo nivel de precios y de empleo. Se puede mostrar que, en este modelo, ambas variables endógenas aumentarán. El dinero no es neutral.

Con respecto a la quinta regularidad, el modelo no falla. El stock de capital es exógeno. En el largo plazo, un aumento en el stock de capital llevará al sistema a un nuevo equilibrio, que implicará un nuevo nivel de precios y de empleo. Se puede mostrar que, en este modelo, el nivel de empleo aumentará y el nivel de precios bajará. Por implicancia, el salario real subirá, pues el salario nominal es fijo. Luego, el nivel de empleo, el nivel de producto y el salario real aumentarán.

Con respecto a la primera regularidad, el modelo falla. De las condiciones del equilibrio general, se deriva la predicción empírica de que el proceso económico operará con desempleo. Si bien el modelo puede explicar la existencia del desempleo, no puede explicar su persistencia. Un aumento significativo en el stock de capital hará que el equilibrio sea con pleno empleo, pues la expansión de la demanda por trabajo

eliminará el exceso de oferta inicial. Una reducción significativa en la oferta laboral, otra variable exógena, hará también que el equilibrio sea con pleno empleo, pues esa reducción eliminará el exceso de oferta laboral. En ambos casos, el mercado laboral se transformará en un mercado walrasiano y el modelo keynesiano en un modelo neoclásico.

Cualquier modelo keynesiano llegará a este mismo resultado con relación a la primera regularidad, pues el equilibrio inicial con desempleo no tiene mecanismos para mantenerse. En realidad, la teoría keynesiana se originó con el propósito de aplicar medidas para llegar al pleno empleo y, así, eliminar el problema social de desempleo. La familia de modelos de la teoría keynesiana es refutada por esta regularidad del capitalismo. Luego, la teoría es refutada por los datos de la realidad y no puede ser aceptada. La sociedad abstracta keynesiana no es una buena aproximación de la sociedad capitalista del mundo real.

Conclusiones

Las tres teorías de la economía estándar fallan en explicar la existencia y persistencia del desempleo, que es una de las regularidades empíricas de los países capitalistas. La teoría neoclásica no explica ni la existencia ni la persistencia. Las teorías clásica y keynesiana explican la existencia inicial del desempleo, pero no pueden explicar su persistencia. En realidad la economía estándar supone, aunque solo de manera implícita, que el desempleo no juega ningún papel en el funcionamiento del capitalismo. Es decir, el capitalismo podría muy bien funcionar con desempleo. Se necesita, por lo tanto, una nueva teoría en donde el desempleo juegue un papel importante en el funcionamiento de la sociedad capitalista.

En el siguiente capítulo se presentarán nuevas teorías que intentan explicar el conjunto de regularidades empíricas del capitalismo, incluyendo la existencia y persistencia del desempleo.

Capítulo 4

La teoría de la inclusión-exclusión

Las teorías de la economía estándar no pueden explicar todas las regularidades empíricas que caracterizan a las economías capitalistas. Se hace necesario entonces examinar nuevas teorías. En el resto del libro presentaremos una teoría que se origina en los recientes trabajos de una nueva corriente teórica que se podría llamar la economía política de la inclusión y exclusión (Figueroa 2003; Nayyar 2003; Sen 1992).

Esta nueva teoría sobre el funcionamiento del sistema capitalista introduce tres nuevos supuestos primarios:

- Existen distintos tipos de sociedades capitalistas, que se diferencian por sus condiciones iniciales, es decir, por su historia; las condiciones iniciales esenciales son dos: la dotación inicial de factores productivos y la desigualdad inicial en la distribución de activos entre los individuos.
- Entre los activos, supone que existen activos económicos (tierra, capital físico y capital humano) y activos sociales (derechos, como el grado de ciudadanía).
- En las diferencias entre sociedades en cuanto a la desigualdad inicial, el factor esencial es la historia colonial de las sociedades.

Si la distribución de activos sociales es desigual, la sociedad es definida como socialmente heterogénea; si es igualitaria, como socialmente homogénea. Por otro lado, si la dotación inicial de factores es tal que la cantidad de capital físico por trabajador es tan baja que da lugar a que la productividad marginal del total de trabajadores sea cero o cercano a cero, la sociedad es definida como superpoblada.

La teoría de la inclusión-exclusión constituye así en un sistema teórico. No será una única teoría que puede explicar el funcionamiento de la sociedad capitalista, como fue el intento de la teoría estándar. Ahora existirán teorías parciales que buscarán explicar el funcionamiento de cada una de las sociedades que componen el sistema capitalista y luego habrá necesidad de una teoría unificada que buscará explicar el sistema como un todo. En ese capítulo se presentarán las teorías parciales y en los capítulos siguientes se construirá la teoría unificada.

Sobre las teorías parciales, consideraremos tres teorías que construirán tres sociedades abstractas. Para puntualizar que se trata de sociedades abstractas, se les dará nombres de letras griegas: épsilon, omega y sigma. Estas sociedades nacieron al capitalismo bajo condiciones iniciales diferentes: épsilon nació socialmente homogénea y no superpoblada; omega nació socialmente homogénea y superpoblada; y, sigma nació superpoblada y socialmente heterogénea.

En este capítulo se estudiarán cada una de estas sociedades con la intención de mostrar que ellas se parecen a las tres sociedades en que se puede dividir el capitalismo del mundo real, según su historia colonial. El primer mundo se parece a la sociedad épsilon; el tercer mundo, que nació al capitalismo sin un legado de dominio colonial, se parece a la sociedad omega; y el tercer mundo, que nació con legado colonial, se parece a la sociedad sigma. Se presentarán modelos de estas sociedades abstractas, cuyas predicciones empíricas serán confrontadas con las regularidades del capitalismo establecidas en el segundo capítulo.

La sociedad épsilon

Las proposiciones alfa de la teoría épsilon incluyen:

Contexto institucional. (a) Reglas: los individuos participan en el proceso económico dotados de activos económicos y sociales; los activos económicos están sujetos a derechos de propiedad; los

individuos llevan a cabo intercambios de mercado, bajo las reglas del mercado, que incluye la regla de que los salarios nominales no pueden bajar. (b) Organizaciones: hogares, firmas y el gobierno.

Condiciones iniciales. Los individuos están dotados con cantidades desiguales de activos económicos, pero existe igualdad en los activos sociales, en el sentido que todos son ciudadanos de primera clase. Existen dos clases sociales: capitalistas y trabajadores. La dotación inicial de factores de la economía —capital por trabajador— es tal que no existe sobrepoblación.

Racionalidad de los agentes. Los individuos actúan guiados por la motivación del interés propio. Los capitalistas buscan objetivos jerarquizados: buscan primero mantenerse en su posición social y solo después buscan maximizar las ganancias. Los capitalistas no alquilan sus bienes de capital en el mercado de alquileres, pues prefieren obtener ganancias a rentas.

Para que esta teoría sea empíricamente refutable, se agregarán los siguientes supuestos auxiliares con el objeto de crear un modelo de la teoría. En la sociedad épsilon se produce un solo bien, el bien B, y existen cuatro mercados: el del bien B, trabajo, dinero y divisas. Los mercados operan bajo competencia perfecta. Los trabajadores están dotados del mismo nivel de capital humano y también de dinero; los capitalistas están dotados de capital físico y también de capital humano. La economía de épsilon es abierta y estática.

Debido a que los trabajadores están dotados de un stock de capital humano homogéneo, existe un único mercado laboral. El mercado laboral opera con un salario nominal que inicialmente está dado, y al cual existe exceso de oferta laboral. Este salario nominal no puede bajar debido a las normas sociales; sin embargo, una reducción en el salario real por la vía de aumentos en el nivel de precio puede ocurrir, pues no hay normas sociales que la impidan.

El desempleo como mecanismo de disciplina laboral

Los trabajadores ofrecen su trabajo a las firmas a cambio de un salario. Se introduce ahora un nuevo supuesto auxiliar sobre el proceso productivo. La productividad laboral depende no solo del stock de capital físico y del número de trabajadores, sino también del esfuerzo de los trabajadores: a mayor esfuerzo, mayor producto total y también mayor producto por trabajador.

Debido a la estructura de clases de la sociedad, los trabajadores no se sienten socios plenos de los capitalistas. Ellos buscan que obtener el mayor salario con el menor esfuerzo posible. El esfuerzo que los trabajadores desplieguen en la firma dependerá de los incentivos que enfrenten. Los capitalistas, por su parte, contratan trabajadores porque así generan ganancias y para ello quisieran el máximo esfuerzo de los trabajadores al salario más bajo posible.

Este conflicto de intereses llevará a los capitalistas a la aplicación de incentivos para obtener el máximo esfuerzo de los trabajadores. El mecanismo que utilizan es la tasa de desempleo. Si no existiera desempleo, si el mercado laboral fuera walrasiano, los trabajadores no tendrían ningún incentivo en poner todo su esfuerzo en la firma, pues si fuesen despedidos por deficiente esfuerzo, igual conseguirían empleo al salario del mercado. En cambio, si existiera desempleo, los trabajadores que no muestren disciplina laboral serían despedidos y sufrirían un costo, el costo del desempleo.

La firma individual estaría interesada en pagar por encima del salario que rige en el mercado, pues así obtendría el máximo esfuerzo laboral y la máxima ganancia. Pero cada firma buscaría hacer lo mismo; en el agregado, las firmas elevarán el salario del mercado por encima del salario walrasiano. En consecuencia, el mercado laboral funcionará con desempleo; será un mercado no walrasiano.

Introducimos ahora otro supuesto auxiliar: existe una tasa de desempleo u^*, que opera como un umbral a partir del cual los trabajadores sienten el costo del desempleo. Esta tasa de desempleo implica un

salario real w* particular, que está por encima del salario walrasiano. Este salario real también opera como un umbral a partir del cual los trabajadores despliegan un alto nivel de esfuerzo y productividad. La intensidad del esfuerzo laboral es, entonces, endógena y depende del salario. Cualquier salario real igual o mayor que w* será denominado *salario de eficiencia*; es decir, a este conjunto de salarios los trabajadores mantendrán la productividad laboral en un alto nivel.

Está en el interés de las firmas que el salario de mercado no sea inferior al salario w*. La solución del mercado laboral está sujeta a esta restricción. Si por alguna razón el salario del mercado fuera inferior a w*, la curva de la productividad laboral se caería y con ello las ganancias. Las firmas tendrían, en consecuencia, todo el incentivo para elevar el salario nominal y restaurar el salario real en el rango del salario de eficiencia.

Equilibrio general

Sobre el proceso económico introducimos ahora otros supuestos auxiliares. La tecnología que utilizan las firmas en la producción del bien B incluye capital físico, capital humano y un insumo material que es importado, el bien C. Esta economía no importa bienes de consumo, solo insumos para la producción. Para pagar por las importaciones del bien C las firmas deben pagar a las firmas extranjeras en una moneda de aceptación internacional, llamada divisa. Para obtener divisas, la firma debe exportar parte de la producción del bien B. Supongamos, además, que el insumo importado no es sustituible por los otros factores. De modo que su uso es en una cantidad fija por trabajador, como si fuese un impuesto en la contratación de trabajadores; por lo tanto, la productividad del trabajo es ahora neta del costo del insumo importado que se requiere para emplearlos. Cuanto mayor es el precio del insumo, la productividad neta del trabajo será menor.

De los cuatro mercados de la economía, tres son walrasianos (bien B, dinero y divisa) y uno es no walrasiano (laboral); de los tres mercados

walrasianos, se puede eliminar uno, apelando a la ley de Walras. El mercado de divisas tiene la particularidad de que, dado el tipo de cambio —el precio nominal de la divisa, digamos, soles por dólar—, el nivel de empleo determina la cantidad demandada de divisas, la cual será necesariamente igual a la cantidad ofrecida, pues ambas decisiones son tomadas por las firmas. El núcleo del equilibrio general se compone, por lo tanto, de solo dos mercados, el mercado de trabajo y mercado de dinero, que deben resolver por las dos variables endógenas el tipo de cambio y el nivel de empleo.

En el mercado de dinero, la demanda viene de los trabajadores y la oferta del gobierno. Al igual que en las economías estándar (ver capítulo 3), los trabajadores mantienen cantidades de dinero en sus bolsillos para hacer frente a las necesidades de transacciones y precauciones que exige el mercado. El saldo real depende positivamente de la masa salarial real: cuánto mayor es el nivel de la masa salarial real, mayor será el requerimiento de saldos reales. Esta relación implica que, para una cantidad dada de la masa salarial real, le corresponderá una cantidad dada de dinero real como inventario; luego, cuanto mayor sea el nivel de precios, mayor tendrá que ser la cantidad de dinero que los trabajadores requerirán para mantener ese inventario real constante.

Epsilón es una economía abierta, que produce un bien B para el mercado interno y para la exportación, cuyo valor en divisas sirve para importar el bien C. Los dos bienes son transables en el mercado internacional; por lo tanto, los precios domésticos no pueden ser independientes de los precios internacionales. Si fuesen diferentes, los individuos comprarían de la fuente más barata y venderían en el lugar más caro. Los precios nominales tienen que igualarse y eso ocurre cuando el precio doméstico es igual al precio internacional multiplicado por el tipo de cambio.

En esta economía, entonces, el nivel de precio depende del tipo de cambio, dado el precio internacional del bien B: a mayor tipo de cambio, mayor nivel de precios. Dado el tipo de cambio, a mayor precio internacional, mayor nivel de precios. La relación entre la demanda de

dinero y el nivel de precios señalada anteriormente se puede ahora establecer con el tipo de cambio: a mayor tipo de cambio, mayor cantidad demandada de dinero nominal.

Como se mostró arriba, el núcleo del equilibrio general está compuesto del mercado laboral y el mercado monetario. La interacción de los dos mercados determinan las dos variables endógenas del núcleo: la cantidad de empleo y el tipo de cambio. La solución se logra de manera simultánea. Las otras variables endógenas se determinan por implicancias. Una vez conocido el nivel de empleo, el nivel del producto queda determinado. Conocido el nivel de precio, el salario real queda determinado. La masa salarial real también queda determinada; y por diferencia con el producto, las ganancias reales también quedan determinadas. La distribución del ingreso nacional queda así resuelta.

La solución de equilibrio en la producción y distribución se puede representar por medio de un gráfico simple. Considere el mercado laboral que aparece en la figura 4.1, donde el salario real se mide en el eje vertical (w) y la cantidad de trabajadores en el horizontal (Q_h). La curva de la productividad marginal bruta está representada por la curva H'N' y la curva de la productividad marginal *neta* por la curva HN, que es la curva de la demanda de trabajo. La curva de oferta laboral es fija al nivel S_h. El salario walrasiano sería el que resulta del cruce entre las curvas de demanda y oferta; pero este no podría ser el de mercado, pues no habría incentivos para la disciplina laboral.

El mercado laboral debe operar con un nivel mínimo de desempleo, que implica un máximo nivel de empleo, dado por el nivel D_h^*. La solución del salario real proviene de las interacciones entre el mercado laboral y el mercado monetario y es igual a w^o. El nivel de empleo de equilibrio es inferior al máximo. El producto total es igual al área que se ubica por debajo del segmento HE y se distribuye entre las ganancias totales P^o y los salarios totales W^o.

El ingreso nacional y su distribución entre los grupos sociales que se muestran en la figura 4.1, así como las otras soluciones del equilibrio general, se pueden escribir como ecuaciones así:

$$Y = P + w D_h$$
$$S_h = D_h + U$$
$$X_b = z^* D_c$$

La primera ecuación muestra que el nivel del producto o ingreso nacional Y se distribuye entre las dos clases sociales, como ganancias P y los salarios totales W. La masa salarial real es igual al salario real w multiplicado por el nivel de empleo D_h. La segunda ecuación dice que la cantidad de trabajadores se divide entre empleados y desempleados; es decir, el equilibrio general es con desempleo (donde U>0). La última ecuación muestra el equilibrio en el sector externo, donde las exportaciones del bien B son suficientes para pagar las importaciones del bien C, donde z^* es el coeficiente de los términos de intercambio internacional (precio internacional del bien B dividido por el precio internacional del bien C).

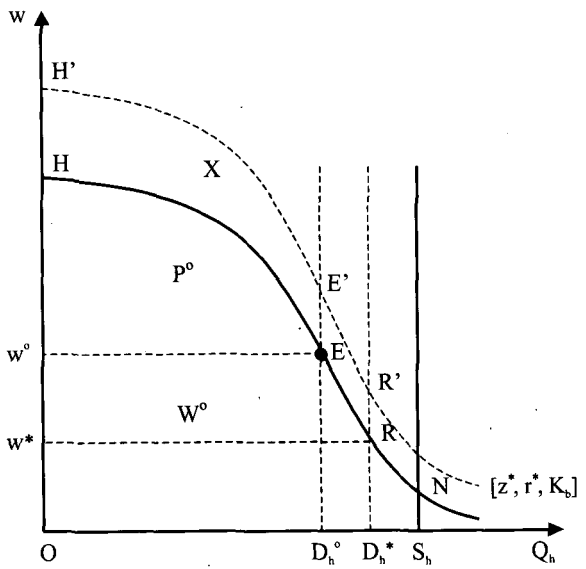

Figura 4.1 Producción y distribución en una sociedad épsilon

Los valores de las variables endógenas que resultan del equilibrio general de la economía se repetirán periodo tras periodo si las variables exógenas se mantienen fijas. La economía es estática porque la solución de un periodo se repite en el periodo siguiente y en los periodos subsiguientes.

Hay que notar que el término «equilibrio» no tiene una connotación ética o deseable; tampoco significa que todos los agentes económicos estén satisfechos con el resultado. Los capitalistas quisieran mayores ganancias de las que han obtenido y los trabajadores quisieran mayores salarios reales de los que han obtenido. El término «equilibrio» quiere decir que no existe agente alguno en el proceso económico que tenga el poder y el deseo de cambiar esta solución.

Predicciones empíricas del modelo y su falsación

Necesitamos ahora generar predicciones empíricas del modelo épsilon para confrontarlo con la realidad. Para ello, se necesita derivar las proposiciones beta: el efecto de las variables exógenas sobre las endógenas. Las variables exógenas son el stock de capital físico, los términos de intercambio internacional, la cantidad de dinero, la cantidad de trabajadores y la desigualdad inicial en los activos.

Un aumento en la cantidad de dinero en la economía tendrá el efecto de aumentar las dos variables endógenas del núcleo: el nivel de empleo y el tipo de cambio. El mayor nivel de empleo implica un mayor nivel del producto. El mayor tipo de cambio implica un mayor nivel de precio y una caída en el salario real. El dinero no es neutral, tiene efectos sobre las variables reales.

Ante aumentos sucesivos de la cantidad de dinero, el empleo aumentará hasta llegar al nivel máximo de empleo, donde la tasa de desempleo es igual al umbral u*. Así se llegará al nivel de pleno empleo *efectivo* —distinto al pleno empleo absoluto donde la tasa de desempleo es cero—. Si la oferta de dinero sigue aumentando, a partir de este punto, todo el efecto se irá a aumentar el tipo de cambio y por lo tanto

al nivel de precio. El dinero se volverá ahora neutral: no afectará las variables reales.

Un aumento en el stock de capital físico hará que la productividad laboral aumente y, por lo tanto, la demanda de trabajo se expandirá. Se generará un exceso de demanda de trabajadores y el equilibrio inicial no podrá persistir. El nuevo equilibrio en las variables endógenas del núcleo será un aumento en el nivel de empleo y una caída en el tipo de cambio. El mayor nivel de empleo implica un mayor nivel del producto; el menor tipo de cambio implica una baja en el nivel de precio y, por consiguiente, un mayor salario real. La masa salarial real aumenta, pero el cambio en el total de ganancias es incierto y, por lo tanto, el cambio en el grado de desigualdad en la distribución del ingreso nacional es también indeterminado.

La relación entre los precios internacionales de los bienes que se exportan sobre los que se importan se denomina *coeficiente de términos de intercambio internacional.* Un aumento en los términos de intercambio internacional, sea por un aumento en el precio internacional del bien B que se exporta o por una baja en el precio internacional del bien C que se importa, tendrá un efecto similar al del stock de capital físico, pues su efecto será el de expandir la curva de demanda de trabajo. A mayores precios de las exportaciones, o menores precios del insumo importado, el valor de la productividad neta del trabajo aumentará.

Por otro lado, un aumento en la cantidad de trabajadores no tendrá efecto alguno sobre el ingreso nacional. Es decir, todo el efecto se irá al mayor desempleo.

Cambios en la desigualdad en los activos se refieren a redistribuciones exógenas en las dotaciones individuales del capital físico —incluidas las tierras—. El capital humano, el otro componente de los activos, no puede redistribuirse. El efecto de una mayor concentración del capital físico no afectará la producción, pues la productividad laboral y por consiguiente la demanda de trabajo no se modificará. La masa de salarios reales no cambiará, tampoco la masa de ganancias. Sin embargo, la desigualdad en la distribución de ingresos aumentará, pues la masa de

ganancias irá ahora a un grupo más pequeño de capitalistas. Hay que notar que en este modelo, donde los capitalistas también están dotados de capital humano, el ingreso de estos incluye salarios por su trabajo, en adición a las ganancias.

El resultado es que la *distribución funcional del ingreso* —entre masas de salarios y ganancias— no cambia con una redistribución del stock de capital físico. Pero la *distribución personal* del ingreso cambia y se hace más desigual. Solo en el caso de la redistribución del capital físico la diferencia entre distribución funcional y personal se hace necesaria. En los demás casos ambos conceptos son iguales.

La matriz de proposiciones beta se presenta en el cuadro 4.1. Las dos primeras variables endógenas se refieren al núcleo del sistema. El resto se obtiene por implicancias. El signo más (+) significa que el efecto de la variable exógena sobre la endógena es positiva, el menos (-) que el efecto es negativo, el cero (0) que no existe efecto alguno y el de interrogación (?) que el efecto es indeterminado. En la columna S_m, que se refiere a la cantidad de dinero, existen dos filas de valores, donde la primera se refiere a situaciones de equilibrio con desempleo y la segunda con pleno empleo efectivo.

¿Se parecen los países del primer mundo y la sociedad épsilon? Las regularidades empíricas relevantes para la confrontación de la teoría con la realidad incluyen las regularidades 1, 4 y 5. El modelo predice la existencia y persistencia del desempleo, la primera regularidad. El modelo también predice que cambios en las variables nominales tienen efectos sobre las variables reales en el corto plazo, la cuarta regularidad. El modelo predice que, en el largo plazo, cuando aumenta el stock de capital físico, aumentan juntos el producto y el salario real, la quinta regularidad. En los tres casos, el modelo no falla.

Los datos de la realidad no refutan el modelo de la teoría épsilon. Por lo tanto, no hay razón alguna para rechazar la teoría en esta etapa de la investigación. La sociedad abstracta épsilon es una buena aproximación del primer mundo.

Cuadro 4.1 Matriz de proposiciones beta del modelo épsilon

Variables endógenas	Variables exógenas				
	S_m	z	K_b	S_h	δ
P_e	+	–	–	0	0
D_h	+,0	+	+	0	0
Y	+,0	+	+	0	0
D	?,0	?	?	+	+
P	+,0	+	?	0	0
W	?,0	+	+	0	0
w	–,0	+	+	0	0
U	–,0	–,0	–,0	+	0

Símbolos: P_e: tipo de cambio nominal; D_h: cantidad demandada de trabajadores; Y: ingreso nacional; D: grado de desigualdad en la distribución del ingreso nacional; P: ganancia total real en las firmas; W: masa salarial real; w: tasa de salario real; U: cantidad de trabajadores desempleados; S_m: cantidad ofrecida de dinero nominal; z: coeficiente de términos del intercambio internacional; K_b: dotación de capital físico; S_h: cantidad ofrecida de trabajadores; δ: grado de desigualdad inicial.

La sociedad omega

La única diferencia entre la sociedad épsilon y la omega es que esta última es superpoblada. Pero para su mayor comprensión, amigo lector, presentamos todo el conjunto de las proposiciones alfa de la teoría omega, que son las siguientes:

Contexto institucional. (a) Reglas: los individuos participan en el proceso económico dotados de activos económicos y sociales; los activos económicos están sujetos a derechos de propiedad; los individuos llevan a cabo intercambios de mercado, bajo las reglas de

este, que incluye como regla que los salarios nominales no pueden bajar. (b) Organizaciones: hogares, firmas y el gobierno.

Condiciones iniciales. Los individuos están dotados con desiguales cantidades de activos económicos, pero existe igualdad en los activos sociales en el sentido de que todos son ciudadanos de primera clase. Existen dos clases sociales: capitalistas y trabajadores. La dotación inicial de factores de la economía —capital por trabajador— es tal que existe superpoblación.

Racionalidad de los agentes. Los individuos actúan guiados por la motivación del interés propio. Los capitalistas buscan objetivos jerarquizados: primero buscan mantenerse en su posición social y solo después buscan que maximizar las ganancias. Los capitalistas no alquilan sus bienes de capital en el mercado de alquileres, pues prefieren obtener ganancias a rentas.

Una economía es superpoblada cuando la dotación inicial de sus factores —capital por trabajador— es tan baja que la productividad marginal laboral es muy baja: cero o cercana a cero. En una economía superpoblada, si se redujera la cantidad de trabajadores en una unidad, el producto total no cambiaría. Considere la sociedad épsilon como referencia inicial, la de la figura 4.1. Suponga ahora una inmigración de trabajadores a épsilon que doble la cantidad de trabajadores. La productividad laboral de toda la población trabajadora —el producto por trabajador— disminuiría notablemente. Debido al principio de los rendimientos decrecientes factoriales, a mayor cantidad de trabajadores, el producto por trabajador disminuye. Así, se llega a un contexto de superpoblación.

¿Podría una sociedad superpoblada funcionar como una sociedad épsilon? Ciertamente, el desempleo resultante sería mucho mayor del que se requiere para el objetivo de asegurar la disciplina laboral, y dejaría de cumplir esa función. Las firmas tendrían que buscar nuevos mecanismos de disciplina laboral. Los trabajadores, por su parte, no podrían

sobrevivir buscando trabajo debido a la baja posibilidad de encontrarlo. La sociedad épsilon no sería viable en un contexto de superpoblación.

Dado que el sector capitalista de omega no podría operar con toda la oferta laboral, su viabilidad depende de que el excedente laboral tenga otras formas de generar sus propios ingresos. Supongamos que los trabajadores tienen la posibilidad de autoemplearse en unidades productivas pequeñas y producir bienes solo con su capital humano —los inmigrantes vienen con la misma dotación de capital humano que los residentes—.

Llamemos sector de subsistencia al conjunto de unidades de producción de autoempleo. «Subsistencia» significa aquí que estas unidades no pueden dar empleo asalariado debido a que la productividad del trabajo es muy baja. Este sector es un sector de refugio de los trabajadores que constituyen el exceso de oferta laboral, que es una proporción importante entre los trabajadores. La economía de omega se compone de dos sectores: el capitalista y el de subsistencia.

Mecanismo de extracción de esfuerzo

¿Cuál sería el mecanismo que utilizarían las firmas para extraer el máximo esfuerzo de los trabajadores en el caso de una sociedad omega?

Supondremos que las firmas establecen el incentivo para la disciplina laboral pagando como salario un monto que está por encima del ingreso que los trabajadores pueden obtener en el sector de subsistencia. El salario estará, por lo tanto, por encima del costo de oportunidad del trabajador asalariado, que es igual al ingreso que se puede obtener en el sector de subsistencia. Luego, el salario incluirá el pago de un premio —digamos 30%— sobre el costo de oportunidad. Si pierde su empleo en la firma, el trabajador sufrirá una pérdida de ingreso que será igual a ese premio. Si el salario fuese igual al costo de oportunidad del trabajador asalariado, perder el empleo no le significaría ningún costo, y entonces no tendría ningún incentivo para trabajar con el mayor esfuerzo en la firma.

Sea w* el mínimo del conjunto de salarios que iguala el costo de oportunidad del trabajador asalariado —su productividad en el sector de subsistencia— más el premio necesario para crearle el incentivo de poner el máximo esfuerzo en la firma. El salario del mercado puede ser mayor a esta cifra, pero no puede ser menor. El salario w* constituye entonces un umbral de salario a partir del cual los salarios mantendrán el nivel de productividad laboral en las firmas; es decir, es el umbral de los salarios de eficiencia. Si el salario de mercado fuese inferior, la productividad laboral y las ganancias se caerían y, por lo tanto, las firmas estarán dispuestas a subir los salarios nominales para restaurar el salario w*. En la sociedad omega, el salario real del mercado no puede ser inferior al umbral de salario de eficiencia. El mercado laboral operará con esta restricción.

Sector capitalista

El modelo de la teoría omega que utilizaremos aquí incluirá un sector capitalista con las mismas características del modelo épsilon. Habrá los mismos cuatro mercados. El núcleo del sistema también será el mismo: los dos mercados, de trabajo y de dinero. Estos dos mercados determinarán de manera simultánea las dos variables endógenas principales del modelo, que son el nivel de empleo asalariado y el tipo de cambio. Esta solución implica una solución de las otras variables: nivel de precios, nivel del producto, salario real y la distribución de ingresos dentro del sector capitalista.

En omega, al igual que en épsilon, la curva de demanda laboral muestra una relación negativa entre el salario real y el nivel de empleo. La solución de equilibrio del salario real y nivel de empleo caerá en un punto de esta curva de demanda laboral, pero sujeto a la restricción de que el salario del mercado no puede ser inferior al umbral de los salarios de eficiencia w*.

La situación de equilibrio en la producción y distribución en el sector capitalista se puede representar por medio de un gráfico simple. Considere la figura 4.2, donde la cantidad ofrecida de trabajadores se

mide en el eje horizontal, y es igual al segmento OO'. Ignore por el momento la cantidad de trabajadores que aparece a continuación de OO'. En el primer eje vertical se mide el salario real y la cantidad de trabajadores empleados en el sector capitalista se mide a partir del origen O.

La demanda de trabajo está dada por la curva HR. Las interacciones entre el mercado laboral y el mercado de dinero dan como resultado la solución en el punto E, con el salario real w^o y el empleo OA. Esta solución debe cumplir con la restricción de que el salario de mercado debe ser un salario de eficiencia. En la figura se cumple con esta condición, pues el salario real de mercado está por encima del umbral w*, que en la figura corresponde al punto g. Si la solución hubiese sido inferior al umbral w*, las firmas habrían elevado el salario nominal hasta llegar a este umbral. La producción y distribución en el sector capitalista también se representa en esta figura a partir del punto E. El exceso de oferta laboral es igual al segmento AO'.

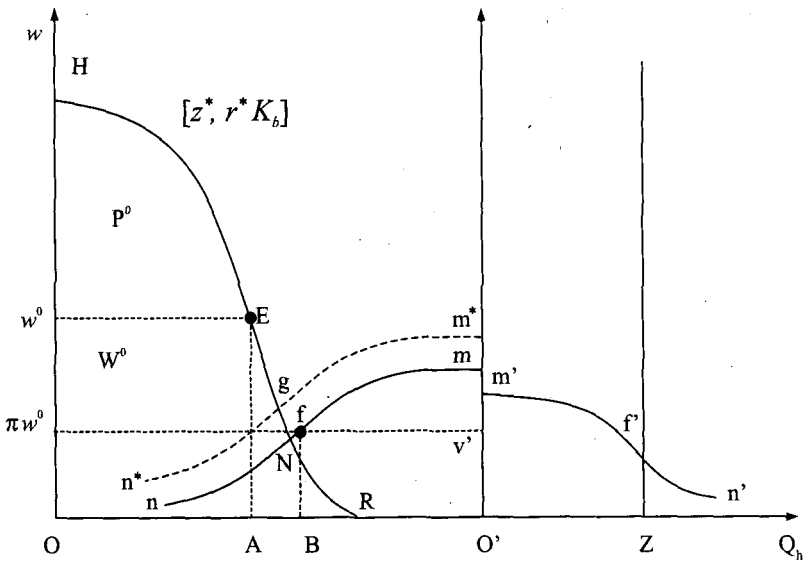

Figura 4.2 Producción y distribución en una sociedad omega (con una cantidad de trabajadores igual a OO')

Sector de subsistencia

Una vez resuelto el equilibrio en el sector capitalista, se determina el exceso de oferta laboral. Un grupo de trabajadores, igual al segmento AO', queda excluido del mercado laboral. Como segunda opción, ellos podrán elegir autoemplearse en el sector de subsistencia o buscar empleo en el mercado laboral.

El sector de subsistencia está representado en la figura 4.2. El autoempleo se mide a partir del origen O' y hacia el origen O. El eje vertical que parte de O' mide el producto por trabajador del sector de subsistencia.

Supondremos que la curva de productividad promedio del trabajo en el sector de subsistencia es decreciente: a medida que aumenta el número de trabajadores disminuye la productividad promedio; es decir, el doble de trabajadores no producirá el doble del producto total, sino menos del doble. Este supuesto implica que la correspondiente curva de productividad marginal es también decreciente y en la figura 4.2 se le representa por la curva mn.

En el sector de subsistencia, el aumento en la cantidad de trabajadores implica el aumento en el autoempleo en unidades productivas pequeñas. Los rendimientos decrecientes que se muestran en la curva mn de la figura 4.2 no se deben a los rendimientos decrecientes factoriales, originados por la utilización de más trabajo a un stock de capital dado, pues en este sector no existen stocks de capital físico. Habrá otra razón detrás de los rendimientos decrecientes. Supondremos que las unidades que se expanden en el sector de subsistencia tienen localizaciones cada vez menos ventajosas, como estar más lejos de los bienes públicos, de los recursos naturales de buena calidad y de los mercados. A este caso se le denomina *rendimientos decrecientes ricardianos*, en honor al economista David Ricardo quien fue el primero en proponerlo.

La curva de la productividad marginal del trabajo en el sector de subsistencia mide el costo de oportunidad de los asalariados y constituye, a la vez, la curva de oferta de trabajo. Es decir, si miramos la curva

mn desde el origen O, desde la perspectiva de las firmas, tendremos la curva nm, que ahora muestra la curva de oferta laboral. Esta curva de oferta es creciente: a mayores salarios reales del mercado, más trabajadores buscarán trabajo asalariado. Como se puede ver en el gráfico, al aumentar el salario real, los trabajadores que al salario inicial obtenían mayor ingreso en el autoempleo y no deseaban buscar trabajo asalariado, estarán ahora dispuestos a ir al trabajo asalariado, pues ahora el salario real es mayor que su ingreso de autoempleo. Por lo tanto, la cantidad de trabajadores que desean empleo asalariado será mayor cuanto mayor sea el salario real.

Sabemos que las firmas están dispuestas a pagar un salario que incluya un premio por encima del costo de oportunidad del asalariado; ahora sabemos que este es igual a la productividad marginal del trabajo en el sector de subsistencia. En consecuencia, la curva relevante para el funcionamiento del mercado laboral no es la curva de la productividad marginal en el sector de subsistencia, sino una curva que se encuentra por encima de esta, pues incluye el premio —digamos 30%—, tal como la curva m*n* en la figura 4.2. Esta es la curva de extracción de esfuerzo.

El salario real walrasiano es el que resulta del cruce de las curvas de oferta y demanda, el punto N en la figura 4.2. Pero esta solución no podría ser de equilibrio, pues no habría incentivos para la disciplina laboral. El mercado laboral no es un mercado walrasiano. El salario real de mercado se determina por las interacciones entre el mercado laboral y el monetario, que es igual a w°, que es un salario de eficiencia, pues se encuentra por encima de la curva de extracción de esfuerzo y asegura entonces la disciplina laboral.

Los trabajadores excluidos pueden entonces elegir entre buscar trabajo en el sector capitalista o autoemplearse en el sector de subsistencia. Y la elección dependerá de cuál rinda más. Para el trabajador individual, el cálculo económico consistirá en comparar el rendimiento de buscar empleo asalariado contra el rendimiento del autoempleo, y elegirá la opción que le produzca más.

Considere el siguiente ejemplo. Buscar empleo asalariado dura seis meses en promedio, entonces el ingreso esperado anual será la mitad del salario total anual. Si en el autoempleo el trabajador puede obtener un ingreso anual mayor, elegirá el autoempleo; si el ingreso en el autoempleo es menor, elegirá buscar empleo asalariado.

Si cada trabajador elige de esta manera, ¿cuál será el resultado en el agregado? Si la productividad laboral es constante, entonces todos los que estén en el autoempleo obtendrán el mismo ingreso. Pero entonces la solución sería de todo o nada: o todos están autoempleados o nadie está autoempleado, dependiendo de cual de las opciones rinda más. No podríamos observar desempleo y autoempleo a la vez, como ocurre en la realidad. Con rendimientos decrecientes, que es el supuesto que hemos adoptado en este modelo omega, en el equilibrio, la curva de la productividad se cortará con el salario esperado —que es una fracción del salario del mercado— y se determinará la cantidad de desempleo y de autoempleo.

El equilibrio en el sector de subsistencia se ilustra en la figura 4.2. Dado que el equilibrio del mercado laboral implica que el exceso de oferta laboral será igual a AO', y dado que el ingreso esperado buscando trabajo es igual a una proporción del salario real, la proporción π, el nivel de autoempleo es igual al segmento O'B. Esta cantidad de trabajadores autoempleados produce el producto total V. El segmento AB es el residuo del residuo y mide el nivel de desempleo.

Como se trata de trabajadores homogéneos en cuanto a su dotación de capital humano, este equilibrio implica que los trabajadores en el autoempleo están *subempleados*. Definimos a un trabajador como subempleado cuando este se encuentra laborando como autoempleado, pero obtiene un ingreso que es inferior al del salario de mercado para una misma calificación. Pueden haber trabajadores autoempleados que pueden ganar por encima del salario del mercado, pero ellos no cumplen con la definición de subempleado. Los supuestos del modelo implican que una proporción importante de los trabajadores autoempleados estará subempleada.

Equilibrio general

El equilibrio general en este modelo omega se determina de manera secuencial. Primero se determina el equilibrio en el sector capitalista. Conocidos el nivel de empleo y el salario real de equilibrio en el mercado laboral, se determina el nivel de subempleo y de manera residual el de desempleo.

Puede ocurrir que el salario esperado es tan bajo que ningún trabajador elige buscar empleo asalariado; es decir, el equilibrio puede ser sin desempleo. El desempleo es doblemente residual y no juega ningún papel en el proceso económico. Es la existencia del subempleo lo que es esencial para el funcionamiento de la economía, pues es esa diferencia de ingresos con el salario lo que asegura la disciplina laboral. El equilibrio general en omega tiene que ser con subempleo; pero el desempleo puede o no existir.

El ingreso nacional y su distribución entre los grupos sociales que aparecen en la figura 4.2 se pueden escribir como ecuaciones de la siguiente manera:

$$Y = Q + V = P + w D_h + v L_s, \text{ tal que } w>v.$$
$$S_h = D_h + L_s + U$$

La primera ecuación muestra que el ingreso nacional Y se compone del ingreso generado en el sector capitalista Q y en el sector de subsistencia V. El ingreso del sector capitalista se distribuye a los capitalistas como ganancias P, a los trabajadores asalariados como masa salarial W —salario real w multiplicado por empleo asalariado D_h—. La masa del ingreso de autoempleo es igual al ingreso por trabajador autoempleado v multiplicado por la cantidad de trabajadores en el sector de subsistencia L_s. La condición de equilibrio es que el salario w sea mayor que el ingreso medio en el autoempleo v.

La segunda ecuación muestra la asignación de los trabajadores. Una parte se encuentra empleada en el sector capitalista, otra está

autoempleada en el sector de subsistencia y otra está desempleada —buscando empleo asalariado y con ingreso cero—.

No hay actor social que tenga el poder y el deseo de cambiar esta situación; es decir, tenemos una situación de equilibrio. Los valores de las variables endógenas se repetirán periodo tras periodo mientras las variables exógenas se mantengan fijas. Cambios en las variables exógenas tendrán efectos sobre las variable endógenas en direcciones determinadas y así podemos obtener las proposiciones beta.

Proposiciones beta y su falsación

Las variables exógenas del modelo omega son el stock de capital, los términos de intercambio internacional, la oferta monetaria, la oferta laboral y la desigualdad en activos. Para que el modelo sea refutable tenemos que obtener ahora las proposiciones beta.

Un aumento en el stock de capital expandirá la curva de demanda de trabajo de las firmas. En el nuevo equilibrio, las dos variables endógenas del núcleo cambiarán así: el nivel de empleo sube y el tipo de cambio baja. Estos resultados implican un mayor nivel de producto, un menor nivel de precio y un mayor salario real. La masa salarial real aumenta pero las ganancias quedan indeterminadas. La consecuencia del nuevo equilibrio en el sector capitalista es que el exceso de oferta laboral baja y el salario esperado aumenta; así, el tamaño del autoempleo disminuye y por lo tanto el ingreso medio aumenta en el sector de subsistencia. El cambio en la distribución del ingreso es indeterminado.

El efecto de los términos de intercambio internacional será similar al del stock de capital. Al aumentar relativamente el precio de internacional del bien B aumentará el valor de la productividad del trabajo y entonces se expandirá la curva de demanda de trabajo.

Un aumento en la oferta monetaria llevará inicialmente a un exceso de oferta en el mercado de dinero. En el nuevo equilibrio, las dos variables endógenas del núcleo, el nivel de empleo y el tipo de cambio, aumentarán. Por implicancia, el nivel de producto aumentará, el nivel

de precio aumentará y el salario real caerá. El exceso de oferta laboral disminuirá y el salario esperado bajará, el autoempleo aumentará, entonces, el ingreso medio caerá. El dinero no es neutral.

Aumentos suficientemente importantes en la cantidad de dinero harán que el nivel de empleo aumente y el salario real disminuya —el movimiento es a lo largo de la curva de la demanda de trabajo— hasta llegar al nivel de pleno empleo efectivo. De esta manera, el mercado laboral llegará al empleo máximo posible y al umbral del salario de eficiencia w^*. Más allá de este punto, inyecciones adicionales de dinero a la economía solo tendrán el efecto de elevar el tipo de cambio y el nivel de precio, sin efectos sobre las variables reales. El dinero devendrá en neutral. La distribución del ingreso en la etapa previa al pleno empleo efectivo será indeterminada. La masa salarial aumenta, pero las ganancias son indeterminadas; el salario real baja pero también lo hace el ingreso medio del autoempleo.

Un aumento en la oferta laboral en una economía sobrepoblada solo tendría, en el corto plazo, el efecto de aumentar el desempleo. Ni el salario esperado ni la curva de rendimientos decrecientes ricardianos se modificarán. En el largo plazo, sin embargo, el umbral del salario real de eficiencia será más bajo —y el nivel de pleno empleo efectivo será mayor—.

Un aumento en la concentración del capital físico no tendrá efecto alguno ni en el nivel del producto ni en la distribución funcional del ingreso. Pero tendrá el efecto de aumentar la desigualdad en la distribución personal del ingreso nacional, pues la misma masa de ganancias irá a un menor número de capitalistas. La matriz de proposiciones beta se presenta en el cuadro 4.2.

Cuadro 4.2. Matriz de proposiciones beta del modelo omega

Variables endógenas	Variables exógenas				
	S_m	z	K_b	S_h	δ
D_h	+,0	+	+	0	0
P_e	+,+	−	−	0	0
Y	+,0	+	+	0	0
Q	+,0	+	+	0	0
V	+,0	−	−	0	0
D	?,0	?	?	+	+
W	?,0	+	+	0	0
P	+,0	?	?	0	0
w	−,0	+	+	0	0
v	+,0	+	+	0	0
L_s	+,0	−	−	0	0
U	−,0	?	?	+	0

Símbolos: P_e: tipo de cambio nominal; D_h: cantidad de trabajadores empleados en el sector capitalista; Y: ingreso nacional; Q: producto total en el sector capitalista; V: producto total en el sector de subsistencia; D: grado de desigualdad en la distribución del ingreso nacional; W: masa salarial real; P: ganancia total real en las firmas; w: tasa de salario real; v: producto por trabajador en el sector de subsistencia; L_s: cantidad de trabajadores en el sector de subsistencia; U: cantidad de trabajadores desempleados; S_m: cantidad ofrecida de dinero nominal; z: coeficiente de términos del intercambio internacional; K_b: dotación de capital físico; S_h: cantidad ofrecida de trabajadores; δ: grado de desigualdad inicial.

¿Se podría decir que los países del tercer mundo y la sociedad omega se parecen? Las proposiciones beta se pueden ahora contrastar contra las regularidades empíricas mencionadas sobre el tercer mundo. Las regularidades relevantes para este modelo estático son 2, 4 y 5. El modelo

omega predice la existencia y persistencia del desempleo y subempleo, la segunda regularidad. El modelo predice relaciones entre variables nominales y reales, como se muestra en la primera columna del cuadro 4.2, la cuarta regularidad. Finalmente, el modelo predice que los salarios reales se mueven en la misma dirección del crecimiento del ingreso nacional de largo plazo, como se puede verificar en la columna 3 del cuadro 4.2, la quinta regularidad.

Los datos conocidos de la realidad no refutan las predicciones del modelo omega. Por lo tanto, no hay razón para rechazar la teoría omega en esta etapa de la investigación. La sociedad abstracta omega constituye una buena aproximación de los países del tercer mundo.

La sociedad sigma

La sociedad sigma muestra una desigualdad inicial también en la distribución de activos sociales. Sigma es socialmente heterogénea. Esta es la única diferencia con la sociedad omega. El conjunto de proposiciones alfa de la teoría sigma es:

Contexto institucional. (a) Reglas: los individuos participan en el proceso económico dotados de activos económicos y sociales; los activos económicos están sujetos a derechos de propiedad; los individuos llevan a cabo intercambios de mercado, bajo las reglas del mercado, que incluye la regla de que los salarios nominales no pueden bajar; sobre los activos sociales, existen normas formales o informales de exclusión de los derechos de ciudadanía. (b) Organizaciones: hogares, firmas y el gobierno.

Condiciones iniciales. Los individuos están dotados con cantidades desiguales de activos económicos; también existe una desigualdad inicial en la distribución de activos sociales, en el sentido de que no todos son ciudadanos de primera clase. Existen dos clases sociales: capitalistas y trabajadores, así como también ciudadanos de distinta

categoría. La dotación inicial de factores de la economía —capital por trabajador— es tal que existe superpoblación.

Racionalidad de los agentes. Los individuos actúan guiados por la motivación del interés propio. Los capitalistas buscan objetivos jerarquizados: primero buscan mantenerse en su posición social y solo después buscan maximizar las ganancias. Los capitalistas no alquilan sus bienes de capital en el mercado de alquileres, pues prefieren obtener ganancias a rentas.

La estructura social de sigma es más compleja que la de omega y épsilon porque se compone de clases sociales y de ciudadanos de distinta categoría. Para establecer un modelo sigma, que haga refutable la teoría, introduciremos varios supuestos auxiliares.

Consideremos que sigma es una sociedad donde la población se puede dividir en tres grupos étnicos o razas: los azules y los rojos como razas primarias y los morados como los mestizos. Las dotaciones iniciales de activos se distribuyen como sigue. Los azules y los morados son ciudadanos de primera clase y ambos conforman las dos clases sociales. Los azules son los capitalistas, con dotaciones de capital físico y capital humano calificado; y los morados son los trabajadores, dotados con capital humano calificado. Los rojos son ciudadanos de segunda categoría y están dotados solo de capital humano que es, además, de bajo nivel (no calificado).

Llamemos grupo social «A» a los azules-capitalistas-ciudadanos de primera; grupo social «Y» a los morados-trabajadores calificados-ciudadanos de primera; y grupo «Z» a los rojos-trabajadores no calificados-ciudadanos de segunda. Como la distribución de los activos económicos y sociales no es aleatoria sino, por el contrario, están muy asociadas o correlacionadas, la sociedad sigma constituye una *sociedad altamente correlacionada*. Se puede entonces establecer una jerarquía en la estructura social, la cual es igual al orden A-Y-Z.

Se puede entender mejor ahora los supuestos que hicimos sobre las sociedades épsilon y omega. Allí existían clases sociales solamente,

pues los activos sociales estaban distribuidos de manera igualitaria. Capitalistas y trabajadores eran ciudadanos de primera categoría. Pudimos haber incluido los tres grupos étnicos y nada hubiera cambiado en el análisis, pues estos tres grupos hubieran sido ciudadanos de primera categoría. En los modelos de épsilon y omega, las categorías sociales y su jerarquía eran solo A-Y. La categoría Z no existe en épsilon ni en omega; existe solo en sigma. Esa es la diferencia social entre las tres sociedades abstractas.

El origen de las clases sociales es fácil de entender. Viene de las diferencias en la dotación del capital físico entre los individuos. La clase capitalista es la que concentra el stock de capital y, por lo tanto, tiene el poder de contratar a los trabajadores y apropiarse de la ganancia. ¿Cuál sería el origen de las diferencias en la ciudadanía? Se origina también en las condiciones iniciales de la sociedad, en su historia. La teoría sigma supone que proviene del legado de dominación social de un sistema colonial o esclavista. El legado de estos sistemas es una sociedad heterogénea y jerarquizada en cuanto a activos sociales y políticos. Los individuos llevan consigo marcas sociales jerarquizadas, dadas por la raza, lengua, religión o costumbres.

Equilibrio general

El modelo de la teoría sigma que vamos a construir será igual al modelo de la teoría omega que presentamos arriba, al cual se le agregará la población Z. Por lo tanto, el estudio de la sociedad sigma se ha simplificado considerablemente. La economía de sigma se compone de tres sectores. Está el sector capitalista con sus cuatro mercados; está también el sector de subsistencia Y, donde se autoemplean los trabajadores Y que han sido excluidos del mercado laboral; y está el sector de subsistencia Z donde se autoemplean todos los trabajadores Z.

El supuesto que estamos haciendo es que el capital humano de los trabajadores Z es muy bajo para la tecnología que utilizan las firmas capitalistas. Por esta razón quedan excluidos totalmente del mercado

laboral. Esta exclusión tiene que ver con su baja dotación de capital humano. Para emplearlos, las firmas tendrían que invertir en su entrenamiento y capacitación; pero eso no es rentable cuando, al mismo tiempo, existe un exceso de oferta laboral de trabajadores calificados, los trabajadores Y.

Los trabajadores Z se autoemplean en unidades pequeñas de producción, donde también producen el bien B. Utilizan tecnología tradicional, distinta a la que se usa en las firmas capitalistas y en las unidades de autoempleo en el sector de subsistencia Y. Por lo tanto la curva de productividad media en el sector de subsistencia Z estará en un nivel por debajo del que corresponde al sector de subsistencia Y, el cual a su vez está por debajo al del sector capitalista. Esta jerarquía en las curvas de productividad laboral entre sectores es el resultado de las diferencias en sus dotaciones de factores.

La curva de la productividad promedio del trabajo en el sector de subsistencia Z muestra también una relación negativa con la cantidad de trabajadores: a mayores cantidades de trabajadores, la productividad media disminuye. La curva correspondiente de productividad marginal es, por lo tanto, también decreciente. La razón es el supuesto de la existencia de los rendimientos decrecientes ricardianos. El bien producido es para el autoconsumo, no para el mercado. El sector Z está inicialmente fuera del mercado de trabajo y del mercado del bien B.

La solución del sector capitalista y del sector de subsistencia Y será exactamente igual a la que vimos en la sociedad omega. Una vez resuelto en los mercados los precios y las cantidades, se hallará la solución en el sector de subsistencia, que ahora llamamos sector Y. Luego, y de manera residual, se determinará el desempleo de los trabajadores Y. Dado que el sector de subsistencia Z está desconectado del sector capitalista, su solución es independiente del resto de la economía. Así llegamos a la solución de equilibrio general de la economía de sigma.

La producción y distribución de equilibrio en la sociedad sigma se puede mostrar por medio de la figura 4.2. Recordemos que la economía de sigma se compone de una economía de omega junto con un sector

de subsistencia Z. La figura 4.2 mostró el equilibrio de una sociedad omega. Solo queda agregar ahora el sector de subsistencia Z en el gráfico para tener la representación de la sociedad sigma. La cantidad de trabajadores Z se mide a partir del origen O' hacia fuera y es igual al segmento O'Z. La curva de productividad marginal laboral del sector está dada por la curva m'n'. El producto total producido en este sector es igual al área que se encuentra por debajo de la curva m'f'.

El ingreso nacional y su distribución entre los grupos sociales en sigma se pueden escribir, utilizando la figura 4.2, en forma de ecuaciones así:

$$Y = P + Q + V_y + V_z$$
$$Y = P + w\,D_{hy} + v_y\,L_y + v_z\,L_z, \text{ tal que } w > v_y > v_z$$
$$S_{hy} = D_{hy} + L_y + U_y$$
$$S_{hz} = L_z$$

La primera ecuación muestra que el ingreso nacional Y se compone del ingreso generado en los tres sectores, capitalista (que se denota por Q), sector de subsistencia Y (que se denota con V_y) y sector de subsistencia Z (que se denota con V_z). Como los tres sectores producen el bien B, la agregación es totalmente válida.

La segunda ecuación muestra la distribución del ingreso nacional entre los cuatro grupos sociales. El producto generado en el sector capitalista se distribuye entre ganancias P, que va a los capitalistas, y la masa salarial, que va a los trabajadores, que a su vez es igual al salario real w multiplicado por la cantidad de trabajadores asalariados D_{hy}. Añadimos luego el ingreso generado en cada sector de subsistencia, descompuesto entre la productividad promedio en cada sector v_y, v_z multiplicado por la cantidad de trabajadores en el autoempleo en cada sector L_y, L_z. Una condición del equilibrio es que el salario real debe ser mayor que el ingreso promedio en el sector de subsistencia Y. Por otro lado, las dotaciones factoriales iniciales llevan a que la productividad media en el sector de subsistencia Y sea mayor que la del sector de subsistencia Z.

La distribución del ingreso muestra diferencias no solo entre capitalistas y trabajadores, sino también entre trabajadores. Los trabajadores Z constituyen el grupo social más pobre de la sociedad. En el medio se ubican los trabajadores Y, excluidos del mercado laboral. Los trabajadores Y que están desempleados no tienen ingresos, pero no son los más pobres. Su elección de desempleo es voluntaria y su ingreso relevante es el salario esperado, que es igual a la productividad marginal del total de trabajadores en el sector de subsistencia Y. Decir que la situación de desempleo es «voluntaria» no es sugerir que el trabajador la considera una situación deseable. Anteriormente se mostró que no lo es, pues esta elección corresponde a una situación de segunda opción, en su condición de trabajador excluido del mercado laboral; su primera opción es ser trabajador asalariado.

La tercera ecuación muestra la asignación de los trabajadores Y al mercado laboral, al autoempleo y al desempleo. La última ecuación dice que la oferta laboral de trabajadores Z se utiliza solo en el autoempleo.

Debido a que ningún actor social tiene el deseo y el poder para modificar este resultado, la situación representada en la figura 4.2, y descrita en las ecuaciones, es de equilibrio. Los valores de las variables endógenas se repetirán periodo tras periodo en tanto las variables exógenas se mantengan fijas. Las proposiciones beta se obtendrán examinando el efecto de cambios en las variables exógenas sobre las variables endógenas.

Proposiciones beta y su falsación

Las variables exógenas del modelo sigma son las mismas que definimos para omega, a las que hay que agregar ahora la oferta laboral de trabajadores Z; las variables endógenas también son las mismas de omega, a la que solo hay que añadir el ingreso medio del sector de subsistencia Z. Las proposiciones beta que se obtuvieron para omega también se aplicarán a sigma. La novedad del modelo sigma está en la desigualdad, pues existe un nuevo grupo social. Las proposiciones beta aparecen en el cuadro 4.3.

Cuadro 4.3 Matriz de proposiciones beta de un modelo sigma

VARIABLES ENDÓGENAS	VARIABLES EXÓGENAS					
	S_m	z	K_b	S_{hy}	S_{hz}	δ
D_{hy}	+,0	+	+	0	0	0
P_e	+,0	–	–	0	0	0
Y	+,0	+	+	0	+	0
Q	+,0	+	+	0	0	0
V_y	+,0	–	–	0	0	0
V_z	0,0	0	0	0	+	0
D	?,0	?	?	+	+	+
w	–,0	+	+	0	0	0
W	?,0	+	+	0	0	0
P	+,0	?	?	0	0	0
v_y	–,0	+	+	0	0	0
v_z	0,0	0	0	0	–	0
L_y	+,0	–	–	0	0	0
U_y	–,0	?	?	+	0	0

Símbolos: P_e: tipo de cambio nominal; D_{hy}: cantidad de trabajadores-y empleados en el sector capitalista; Y: ingreso nacional; Q: producto total en el sector capitalista; V_y: producto total en el sector de subsistencia-y; V_z: producto total en el sector de subsistencia-z; D: grado de desigualdad en la distribución del ingreso nacional; w: tasa de salario real; W: masa de salario real; P: ganancia total real en las firmas; v_y: producto por trabajador en el sector de subsistencia-y; vz: producto por trabajador en el sector de subsistencia-z; U_y: cantidad de trabajadores-y desempleados; S_m: cantidad ofrecida de dinero nominal; K_b: dotación de capital físico; S_{hy}: cantidad ofrecida de trabajadores-y; S_{hz}: cantidad de trabajadores-z; δ: grado de desigualdad inicial.

Hemos definido arriba dos tipos de países en el tercer mundo, según sus condiciones iniciales o su historia: países que nacieron al capitalismo sin un legado colonial —o con un legado débil— y aquellos que nacieron con un legado colonial significativo. Estos últimos son los que la teoría sigma intenta explicar. La población Z se compone de los descendientes de las poblaciones que fueron las dominadas en el sistema colonial, poblaciones aborígenes y esclavizadas. Por lo tanto, las regularidades empíricas relevantes para confrontar con el modelo sigma, un modelo estático, son las regularidades 2, 3, 4 y 5, según fue establecido en el segundo capítulo.

El modelo sigma predice la existencia y persistencia del desempleo y subempleo, la segunda regularidad. El modelo sigma también predice que la población Z —descendientes de la población dominada en el sistema colonial o en el sistema esclavista— es la más pobre de la sociedad, la tercera regularidad. El modelo sigma predice que el dinero no es neutral en el corto plazo, la cuarta regularidad. Finalmente, el modelo sigma predice que el nivel del producto y el salario real aumentan en el largo plazo, cuando el stock de capital aumenta: la quinta regularidad.

En suma, los datos de la realidad no refutan las predicciones del modelo sigma. No hay, por lo tanto, razón alguna para rechazar la teoría sigma en esta etapa de la investigación. La sociedad abstracta sigma es una buena aproximación de los países del tercer mundo que tienen un legado de dominación colonial importante. Por su parte, la sociedad abstracta omega es una buena aproximación de los países del tercer mundo que no tienen un legado de dominación colonial.

La mayoría de los países del primer mundo se iniciaron como países capitalistas sin un legado de dominación colonial importante; más bien, ellos fueron los colonialistas. Este es el caso de los países de Europa occidental y de Japón. También es el caso de los países que surgieron al capitalismo luego de un periodo colonial, de relativamente corto periodo y en territorios relativamente poco poblados, los sistemas de colonizadores *settler colonies*, tales como Estados Unidos, Canadá y

Australia. En general, los países del primer mundo pasaron de sociedades tipo omega a sociedades tipo épsilon.

La historia del tercer mundo es diferente: la mayoría tiene una historia de conquista y de dominación colonial *non-settler colonies*. Los países del tercer mundo que no tienen esta historia, o lo fueron solo por periodo breve —menos de cincuenta años— son pocos. Argentina, Corea del Sur, Israel, Tailandia, Taiwán y Uruguay constituyen los ejemplos notables de sociedades tipo omega. La teoría sigma es, por lo tanto, mucho más relevante para explicar el funcionamiento del tercer mundo que la teoría omega.

Desigualdad y estructura productiva

En los países del tercer mundo se observa una gran cantidad de pequeñas unidades productivas, tanto en el campo como en la ciudad. Estas unidades corresponden a los sectores de subsistencia de la teoría sigma. La explicación que da la teoría es que todo el empleo que se observa en esos sectores es una medida del exceso de oferta laboral de equilibrio.

El modelo sigma supone que la producción de los sectores de subsistencia es solo para el autoconsumo y no para el mercado. Si fuese también para el mercado, ¿de dónde vendría la demanda? Como los sectores de subsistencia producen con una dotación de capital muy baja o nula, se puede suponer que producen bienes distintos a los que produce el sector capitalista; puntualmente, que producen bienes inferiores, bienes de baja calidad. Luego, la demanda vendría de los pobres, es decir, de los mismos trabajadores de los sectores de subsistencia. Los pobres producen para los pobres. Así se forma una subeconomía de los sectores de subsistencia.

El modelo sigma supone precisamente esta situación. Aunque producen también para el mercado, es como si los sectores de subsistencia produjeran solo para el autoconsumo. Según el modelo sigma, el factor esencial que explica la existencia de los sectores de subsistencia está en

la dotación factorial inicial. Los mercados de bienes también tienen un efecto, pero este efecto es pequeño y se puede ignorar.

Si se introduce mercados de bienes en los sectores de subsistencia, la solución de equilibrio general será como sigue. Dada la desigualdad inicial en la distribución de los activos, el equilibrio general en la economía de sigma opera con un gran exceso de oferta laboral y de una gran desigualdad en la distribución del flujo del ingreso nacional. También incluye una estructura productiva desigual en dotaciones factoriales que producen bienes de distinta calidad para mercados diferenciados que resultan de la desigualdad en los ingresos. Los pobres producen principalmente para los pobres. La economía de sigma es un todo integrado y consistente; es decir, opera en una situación de equilibrio general.

Esta explicación es diferente a otra muy popular que atribuye la existencia de las unidades productivas pequeñas a la política de los gobiernos que encarecen el costo de hacerse legales y atribuyen a las pequeñas unidades el carácter de actividades ilegales. En la literatura internacional no existe una teoría formal sobre la «economía ilegal o informal» de la cual se hayan derivado proposiciones refutables. Tampoco existen, ni pueden existir, trabajos empíricos que hayan hecho tal refutación, que es lo que se necesita para que sea una explicación alternativa.

Según la teoría omega o sigma, aunque el costo de hacerse legal fuese cero, los países del tercer mundo funcionarían con sectores de autoempleo en pequeñas unidades. El factor esencial no está en los costos de la legalidad, ni en ningún otro factor, sino en las condiciones iniciales de esas economías.

Conclusiones

Hemos considerado tres sociedades capitalistas abstractas con la finalidad de que expliquen el comportamiento del primer mundo, del tercer mundo con un legado de dominación colonial significativa y el tercer

mundo sin ese legado. Estas realidades, en efecto, se parecen a las sociedades abstractas épsilon, sigma y omega. Las predicciones de cada modelo construido no han sido refutadas por las regularidades empíricas de estos grupos de países, las regularidades 1 al 5 que se establecieron en el segundo capítulo.

Hemos logrado así construir tres buenas teorías parciales del sistema capitalista, «buenas» en el sentido que no han podido ser refutadas por los datos de la realidad. Las cinco regularidades empíricas de los países capitalistas son explicadas por estas teorías.

A este conjunto de teorías le hemos llamado teorías de la inclusión y exclusión. En efecto, estas teorías muestran que el proceso económico en una sociedad capitalista opera con mecanismos de inclusión y exclusión de los trabajadores a la sociedad a través del mercado laboral. Los distintos tipos de sociedades capitalistas operan con distintos mecanismos y dan lugar a diferentes soluciones, tanto en el nivel de ingreso como en el grado de desigualdad.

Falta explicar las dos regularidades empíricas —6 y 7—, que se refieren al conjunto del sistema capitalista. El sistema capitalista real se compone, según nuestra clasificación, del primer mundo y del tercer mundo, dividido a su vez en dos. En términos teóricos, el sistema capitalista abstracto se compone de tres sociedades abstractas, épsilon, omega y sigma, que explican el funcionamiento de cada parte del capitalismo. Hemos llegado, hasta aquí, a construir buenas teorías del capitalismo, pero son *teorías parciales*. Las regularidades que faltan explicar necesitan una *teoría unificada* del capitalismo, que pueda explicar el sistema capitalista considerado como un todo.

Surge ahora la pregunta, ¿buenas teorías parciales implican una buena teoría unificada? No necesariamente. La física nos ha mostrado este problema. Dos teorías parciales, una que explica las relaciones entre objetos en el mundo de lo pequeño o subatómico —la teoría cuántica— y la otra que explica las relaciones en el mundo de lo grande —la teoría de la relatividad— son, sin embargo, inconsistentes entre sí. Las dos teorías no pueden ser ciertas. El actual desafío

en la física es desarrollar una teoría unificada, la *teoría del todo*, que explique ambos mundos.

De manera similar, se puede decir que en la ciencia económica tenemos ahora tres buenas teorías parciales, pero eso no asegura que tengamos una teoría unificada. Esas teorías parciales pudieran ser inconsistentes entre sí y no poder explicar el sistema capitalista como un todo. También en la ciencia económica necesitamos desarrollar una teoría del todo.

Según las proposiciones alfa de las teorías, las sociedades capitalistas se diferencian por sus condiciones iniciales, por su historia. Sigma es una sociedad que nació al capitalismo con una diversidad social —como multiétnica, multicultural y multilingüe— que la convirtió en una sociedad heterogénea y jerarquizada en cuanto a activos sociales y políticos. Omega y épsilon nacieron al capitalismo como sociedades homogéneas. Pudieron haber nacido al capitalismo también con esa diversidad social, pero esta no las condujo a una sociedad jerárquica, pues no tienen un legado de dominación colonial.

Estos son los supuestos primarios de las tres teorías parciales. En los tres casos, la pregunta es si sus condiciones iniciales cambiarán o no cambiarán endógenamente en el tiempo. Esta es una de las preguntas fundamentales de este libro. Su respuesta requiere de una teoría unificada y de modelos dinámicos. Estos serán desarrollados en el resto del libro.

Como preparación a la presentación de la teoría unificada, desarrollaremos algunas teorías intermedias. Los tres capítulos siguientes tienen ese objetivo.

Capítulo 5

Desigualdad y desorden social

El capítulo anterior mostró el funcionamiento de tres tipos de sociedades capitalistas. El equilibrio general en cada sociedad fue presentada. En particular, se mostró que «equilibrio general» no quiere decir que todos los individuos de la sociedad están satisfechos y felices con la solución de producción y distribución. Todos quisieran más bienes, pero dada las restricciones individuales y agregadas, ningún actor tiene el poder y el deseo para cambiar la solución. Equilibrio significa una situación que no puede ser modificada endógenamente.

La pregunta ahora es, ¿equilibrio general implica orden social?, ¿podría existir equilibrio general con desorden social? En este libro, el término *desorden social* se referirá al comportamiento de la gente que va en contra de las reglas del contexto institucional.

¿Cuál es el origen del desorden social? Como ya sabemos, para dar respuesta a este tipo de preguntas se necesita una teoría; y hacer teoría implica establecer un conjunto de supuestos primarios. En este capítulo pues, se presentará una teoría que intente explicar el desorden social.

El primer paso es suponer que el orden social es un bien, que la gente prefiere tenerlo a no tenerlo, o que prefiere un grado de mayor orden social a otro de menor grado; además, que el orden social es un *bien público*, un bien que solo se puede producir y consumir de manera colectiva. No podemos entrar a una bodega o supermercado y comprar unas cuantas unidades de orden social para consumirlo en

casa. O todos disfrutamos del orden social o todos nos perjudicamos del desorden social.

¿Cuál es el origen del desorden social? ¿Cuál es el papel de la desigualdad? Una teoría que intenta explicar ese papel se presenta a continuación.

La teoría de la tolerancia limitada a la desigualdad

Así como los individuos prefieren un ingreso mayor a otro menor, supondremos que también prefieren un ingreso relativo mayor —relativo al ingreso promedio de su grupo de referencia— a otro menor. Los individuos prefieren tener una posición económica mayor a otra menor. La desigualdad entra en sus preferencias.

Los individuos aceptarán la desigualdad dentro del grupo social pero solo hasta cierto grado; la desigualdad que va más allá les causará resentimiento o envidia y será considerada inaceptable. Supondremos que la gente tiene una tolerancia limitada a la desigualdad. Las reacciones ante una desigualdad considerada excesiva implicarán romper alguna regla del contexto institucional, pues los individuos considerarán que el sistema es injusto. Si esas reacciones involucran a un individuo solamente, o a pocos, la desigualdad no tendrá mayor consecuencia para el funcionamiento de la sociedad —el individuo tendrá un problema—; pero, si involucra a muchos individuos se generará el desorden social —la sociedad tendrá ahora un problema—. Cuanto mayor sea el número de individuos que consideren que la desigualdad que existe en la sociedad es inaceptable, mayor será el grado de desorden social.

Se puede entonces construir una teoría sobre el comportamiento de los individuos frente a la desigualdad en la sociedad. La proposición alfa, que intenta ser válida para todo tipo de sociedad capitalista —épsilon, omega o sigma—, se puede expresar así:

Existe en los individuos una tolerancia limitada a la desigualdad en la sociedad. Los individuos tienen un sentido de justicia distributiva que los lleva a tener umbrales de tolerancia a la desigualdad que existe en la sociedad. Si la desigualdad sobrepasa esos umbrales, los individuos reaccionarán y buscarán restaurar la desigualdad a los niveles tolerables.

Para obtener un modelo de esta teoría, introduciremos el supuesto auxiliar de que los individuos tienen distintos umbrales de tolerancia. Para un grado de desigualdad dado habrá un conjunto de situaciones sobre tolerancia. La primera es que todos toleran esta desigualdad. Este es el caso de una desigualdad que es socialmente tolerada. Esta desigualdad no causa desorden social.

La segunda es que un grupo la tolera pero otro grupo no lo hace. Esta desigualdad no es socialmente tolerada y dará origen al desorden social. Si el grado de desigualdad aumenta, el grupo que no toleraba la desigualdad anterior tampoco tolerará esta que es mayor; pero del grupo que toleraba la desigualdad anterior, habrá ahora una parte que no tolere la desigualdad actual. Por lo tanto, habrá más gente que no tolera un mayor grado de desigualdad. La consecuencia para el agregado es que a mayor desigualdad le corresponderá un mayor grado de desorden social.

La proposición beta que se deriva de este modelo dice que deberíamos observar en la realidad una relación positiva entre desigualdad y desorden social: a mayor grado de desigualdad le deberá corresponder un mayor grado de desorden social. Pero esta relación es apenas una de varias que se dan en la sociedad y para encontrar la situación de equilibrio en el grado de desorden social se necesita introducir las acciones de otros grupos sociales y luego las interacciones entre ellos. Ciertamente, el comportamiento del gobierno frente al desorden social será el otro componente de la teoría.

Comportamiento de los gobiernos

¿Cuál es el papel que juegan los gobiernos en el proceso económico? Un papel importante es la provisión de bienes públicos.

Existen bienes que no pueden ser producidos y consumidos de manera privada. No todos los bienes de la sociedad son bienes privados. También existen los llamados *bienes públicos*, cuyo consumo es colectivo, pues nadie puede ser excluido de su consumo. La producción de bienes públicos enfrenta varios tipos de limitaciones. Los bienes públicos relevantes para nuestro análisis son indivisibles (como las carreteras). No podrían, por lo tanto, ser producidos por las firmas porque no podrían ser vendidos en el mercado a consumidores individuales. Tampoco pueden ser producidos por los propios consumidores a través de una acción colectiva debido al llamado *problema olsoniano,* en honor al economista Mancur Olson, quien fue el primero en formularlo. El problema es el siguiente: si la gente actúa guiada por su interés personal, un bien público no será producido. Nadie deseará pagar los costos de producir un bien público, pues una vez producido, todos se beneficiarán, tanto los que pagaron como los que no pagaron. El incentivo individual es entonces no pagar, esperando que los otros lo hagan. Como todos los individuos actuarán de la misma manera, en el agregado nadie estará dispuesto a pagar y el bien público no se podrá producir.

Frente a estas dificultades de la iniciativa privada, quedan solo los gobiernos como los actores sociales que pueden producir los bienes públicos. Ellos pueden utilizar mecanismos coactivos para producirlos, como la aplicación de impuestos.

El dinero es un ejemplo de bien público. Cada uno no podría producir su propio dinero. El dinero tiene que ser producido y utilizado colectivamente. Una vez que un bien se convierte en dinero, en medio de pago, nadie puede ser excluido de utilizarlo y beneficiarse de su uso. Los servicios que presta la infraestructura social también son bienes públicos. Cada uno no podría producir su propia carretera. Las carreteras son de uso colectivo. Los pobladores de una región que se beneficiarán

de la carretera tampoco tendrán incentivos para producirla por una acción colectiva. El gobierno local o regional resolverá este problema olsoniano coaccionando a los pobladores el pago del costo de la carretera mediante la aplicación de un impuesto.

En suma, allí donde se necesita producir un bien público que la acción colectiva de los individuos no puede producirlo, se necesita la acción del gobierno. Cuando la acción colectiva falla, cuando la racionalidad individual falla, el bien público se puede producir de manera coercitiva. Eso necesita de la autoridad del Estado. En realidad, son los bienes públicos los que contribuyen a que la sociedad exista, a que la sociedad sea más que la suma de los individuos, pues induce a la existencia del Estado y de los gobiernos. Con la provisión de un bien público todos los miembros de la sociedad estarán bien; si este bien no existe, todos estarán mal.

La racionalidad de los gobiernos

El orden social es un bien público, como se dijo anteriormente. Como no hay incentivos privados para producirlo por medio de la acción colectiva, el gobierno es el actor social que puede producirlo. El desorden social implica que las reglas de juego no se respetan. La consecuencia del desorden social es, por lo tanto, que la sociedad funcionará con mayor violencia, ilegalidad, corrupción, desconfianza, burocratismo. Los costos de funcionar como sociedad, de vivir la vida cotidiana, llamado los costos de transacción, serán más altos para todos.

El desorden social también desafiará la legitimidad del sistema político y económico, así como también la del gobierno. Llevará, por lo tanto, a la inestabilidad política. Los gobiernos pueden o acabar su periodo antes de tiempo o perder las próximas elecciones. El sistema democrático puede interrumpirse. Frente a estos posibles efectos sobre el proceso político, los gobiernos tendrán incentivos para evitar el desorden social o controlarlo.

¿Cuál es el origen del desorden social? Según la teoría de la tolerancia limitada a la desigualdad, mostrada anteriormente, la excesiva desigualdad causará desorden social. Este es el costo social de la desigualdad. Luego, los gobiernos tendrán que actuar también sobre la desigualdad para controlar el desorden social. Los instrumentos a utilizar para controlar el desorden social incluyen la represión y la redistribución. Necesitamos una teoría de los gobiernos para explicar su comportamiento frente a la desigualdad.

El supuesto de que las acciones del gobierno se determinan exógenamente es muy común en la literatura económica. Los gobiernos, en esta literatura, pueden elegir libremente las variables nominales, así como los componentes del presupuesto público, tales como nivel de gasto, impuesto, déficit endeudamiento. Este supuesto está tan extendido en la literatura económica que es usual encontrar trabajos que concluyen con recomendaciones sobre la política que deben seguir los gobiernos, en donde los autores les dicen a los gobiernos lo que debe hacer, como si los gobiernos no tuvieran sus propios intereses.

Recientemente, ha aparecido dentro de la escuela neoclásica una nueva teoría —llamada la «elección pública»— que considera que los gobiernos tienen sus propias motivaciones y que su comportamiento está entonces endógenamente determinado. La idea básica es que la gente que está en el gobierno es como cualquier otra gente, busca sus propios intereses. Las acciones de los políticos podrían también estar guiadas por otros intereses, como su ideología, o el altruismo; pero se supone que estos factores no son los más importantes. El supuesto de esta teoría es que la motivación esencial de los individuos que participan en el gobierno está en la búsqueda del propio interés y que cualquier efecto positivo sobre el bienestar social aparecerá como un subproducto de esta racionalidad.

Aquí adoptaremos este supuesto y agregaremos otros para presentar una teoría de los gobiernos. La teoría se puede expresar como una proposición alfa, que intenta ser válida en toda sociedad, así:

Racionalidad de los políticos. Los políticos buscan dos objetivos que están jerarquizados: primero la permanencia en la clase política y luego la maximización del ingreso.

Esta teoría establece la existencia de una clase social: la clase política. Al igual que en la racionalidad de los capitalistas, los políticos buscan mantenerse en la clase política y buscan maximizar sus ingresos, donde el primer objetivo tiene prioridad. El objetivo de maximizar los ingresos no puede comprometer la posición social, es decir, los políticos no están dispuestos a perder poder político a cambio de mayores ingresos. Sus ingresos de largo plazo están asegurados manteniéndose dentro de la clase política.

De manera también similar a los capitalistas, la teoría de los gobiernos tiene la implicancia de que el comportamiento de los gobiernos es endógeno. Los gobiernos interactúan con el resto de los actores sociales de la sociedad y no están por encima de ellos.

Se supone que esta teoría es, en principio, válida para todo tipo de sociedad capitalista. También es válida, en principio, para gobiernos democráticos y no democráticos. El capitalismo puede funcionar con ambos tipos de sistemas políticos; y si el capitalismo es democrático, puede funcionar con distintos grados de democracia, que van desde las formas representativas hasta las participativas. El mercado y la democracia son las dos instituciones básicas del capitalismo, pero como veremos más adelante, y al igual que el grado de desarrollo de los mercados, el grado de la democracia en el capitalismo —desde la no democracia hasta las formas de democracia participativa— es una variable endógena.

Para llevar esta teoría a la falsación, necesitamos construir un modelo. Introduciremos supuestos auxiliares para tal efecto. Dejaremos de lado el caso de los gobiernos no democráticos. Como contexto en el cual operan los gobiernos, supondremos las tres sociedades capitalistas —épsilon, omega y sigma— de nuestro estudio; sin embargo, por simplicidad, ignoraremos la sociedad omega, que para el proceso político es similar a épsilon, pues es también socialmente homogénea.

El modelo supondrá que los gobiernos democráticos buscan maximizar votos, pues los votos les dan la legitimidad al poder político, sujeto a las restricciones del presupuesto público, de las normas institucionales y de las demandas de los grupos de presión. La distribución del poder entre los grupos sociales para ejercitar sus demandas está determinada por la distribución de los activos económicos y políticos entre los grupos sociales.

Dada las variables exógenas que enfrentan, los gobiernos buscarán alcanzar su mejor posición, la de equilibrio, en cuanto a legitimidad política a través de los votos —y a la popularidad que también es una forma de votación—. Este objetivo tiene varias implicancias empíricas sobre el comportamiento de los gobiernos.

Una primera implicancia empírica es que los gobiernos buscarán utilizar el presupuesto público para comprar votos; es decir, los gobiernos buscarán que los gastos fiscales sean mayormente discrecionales —con los cuales pueden comprar votos—, en lugar de tener que hacer gastos obligatorios para financiar los derechos ya establecidos de los ciudadanos —con los cuales no pueden comparar votos—.

La segunda implicancia es que los gobiernos buscarán utilizar el endeudamiento antes que la recaudación de impuestos para financiar el gasto público; la primera forma de financiamiento no tendrá efecto en las próximas elecciones, pero la segunda tendrá un efecto negativo. La tercera implicancia es que los gobiernos buscarán utilizar la política fiscal de acuerdo al ciclo político; es decir, aumentarán el gasto antes de las elecciones y lo reducirán después de las elecciones. El gasto fiscal seguirá al ciclo político. La cuarta implicancia es que los gobiernos buscarán asignar el gasto fiscal entre sectores y regiones de manera, así maximizarán la rentabilidad política; por ejemplo, el gasto será mayor en las grandes ciudades que en las áreas rurales, en obras que se pueden inaugurar, antes que en innovaciones institucionales.

En general, los gobiernos tendrán un comportamiento miope. Los efectos de corto plazo de la política fiscal tendrán mayor importancia para su objetivo de comprar votos que los de largo plazo.

Estas predicciones empíricas del modelo no han sido objeto de estudio en la literatura internacional. Sobre la existencia del ciclo político en el gasto público, un estudio encontró que, en efecto, en los Estados Unidos el gasto social del gobierno aumenta en periodos previos a elecciones y disminuye después de estas (Rogoff 1990).

Comportamiento frente a la desigualdad

¿Cuál es el comportamiento del gobierno frente a la desigualdad que genera el mercado? Hay que notar que en las tres sociedades capitalistas la desigualdad resulta del funcionamiento del mercado. Esto es claro en el caso de la sociedad épsilon. En omega y sigma, donde existen sectores de subsistencia, el equilibrio general es secuencial, donde el producto y el empleo se determinan primero en el mercado y por residuo en los sectores de subsistencia. Y de manera similar, la desigualdad que emerge del mercado determina la desigualdad en el conjunto de la sociedad.

Considere el caso en el cual la distribución del ingreso que emerge del mercado implica una excesiva desigualdad y genera desorden social. En el objetivo de reducir el desorden social, que le afecta los votos, los gobiernos cuentan con dos instrumentos: la represión y la redistribución. La medida de la represión tiene efectos inmediatos en poner orden en la sociedad, mientras que la redistribución tendrá resultados en periodos largos y no será políticamente rentable en las próximas elecciones. Además, los grupos de presión más importantes son los capitalistas y ellos se opondrán a una redistribución significativa de ingresos. La desigualdad no se modificará.

Esta condición de equilibrio puede parecer contraintuitiva. En un sistema democrático se espera que las decisiones colectivas se tomen por la regla de la mayoría, la cual a su vez implica que el votante del centro es el que dirime. Este es el llamado *principio del votante mediano*. En una sociedad desigual, el votante mediano tendrá un ingreso por debajo del ingreso medio; por lo tanto, el votante dirimente pertenecerá al grupo de los pobres y votará a favor de los pobres. En términos

estadísticos, en una distribución del ingreso que sea normal, en forma de campana, simétrica, donde por cada rico existe un pobre, el valor de la media y la mediana son iguales; por contraste, en una distribución muy desigual, que es asimétrica, donde por cada rico existen muchos pobres, la mediana es menor que la media.

En consecuencia, uno esperaría que en un sistema democrático, el principio del votante mediano lleve naturalmente a los gobiernos a aplicar medidas redistributivas hasta llegar a una desigualdad socialmente aceptada. Si fuese así, la excesiva desigualdad y el consecuente desorden social se autorregularían. Es más probable que este comportamiento ocurra en una democracia participativa. Pero en una democracia representativa, que es la más común en la realidad, los votantes no eligen las políticas públicas, sino son sus representantes los que eligen esas políticas, y lo hacen guiados por sus propios intereses.

En las relaciones entre gobernantes y gobernados existe un problema de agente-principal, donde el principal lo conforman los votantes y el agente es el gobierno. Se denomina *problema de agente-principal* al que aparece cuando el agente tiene incentivos que no son compatibles con los objetivos que persigue el principal. Este es también el caso en el mercado laboral, en las relaciones entre capitalistas y trabajadores.

La motivación de la búsqueda de la maximización de los votos no conduce a los gobiernos a la reducción de la excesiva desigualdad que emerge del funcionamiento del mercado y que genera desorden social. Los gobiernos no tienen ni el poder ni el incentivo para llevar la desigualdad al rango de la tolerancia social. Pero los gobiernos tampoco tienen incentivos para aumentar la desigualdad que emerge del mercado, cualquiera que esta sea, pues implicaría aumentar el grado de desorden social.

Se puede distinguir dos tipos de desigualdad. La desigualdad que emerge del mercado se puede denominar la *distribución primaria*. Las acciones del gobierno para modificarla, así como sus interacciones con los otros actores sociales, darán lugar a una *distribución secundaria*. El modelo teórico propuesto aquí predice que estas dos distribuciones no

serán muy distintas. Los gobiernos no tienen ni el poder ni los incentivos para modificar la distribución primaria. Los grupos de presión con poder son los mismos que concentran la distribución de activos económicos y políticos de la sociedad. Los gobiernos pueden llevar a cabo programas a favor de los más pobres y así comprar votos, pero eso solo conduce a mantener la desigualdad primaria, pues los pobres también pagan impuestos y los ingresos de los ricos pueden también aumentar. El equilibrio distributivo puede no ser socialmente aceptado, lo que podría generar desorden social.

En suma, los gobiernos no buscan reducir la desigualdad que emerge del mercado. Esa no es su motivación. Su motivación es mantenerse en el poder y para ello buscan comprar votos. Cualquier acción del gobierno para reducir la desigualdad será un subproducto de su motivación. A diferencia de los trabajadores y capitalistas, los gobiernos no tienen un umbral de tolerancia a la desigualdad. La situación de equilibrio distributivo implica que la distribución primaria y la secundaria no serán muy distintas. La otra predicción empírica del modelo es que la situación de equilibrio en el comportamiento de los gobiernos puede darse con cualquier grado de desigualdad, incluido el que genera desorden social. La tercera es que los gobiernos le darán mayor prioridad a los programas a favor de los pobres que a cambiar la desigualdad. Estas predicciones se derivan del comportamiento de los gobiernos en su búsqueda de la mejor posición, considerando que las variables exógenas están dadas, y son todas empíricamente refutables.

Cambios en las variables exógenas modificarán la situación de equilibrio y así el comportamiento de los gobiernos. Una de estas variables es la desigualdad en la distribución de los activos económicos y políticos entre los grupos sociales, lo que implica una distribución del poder entre los grupos de presión. Si esta distribución cambiara, el comportamiento de los gobiernos también se modificaría.

En lugar de analizar el efecto de cambios en la distribución de activos en una sociedad dada, podemos ver este efecto en sociedades capitalistas que se diferencian en esa distribución. Consideremos los dos tipos de

sociedades capitalistas abstractas épsilon y sigma. En la sociedad épsilon, que es socialmente homogénea, donde la ciudadanía es uniforme, el contrato social incluirá la norma de establecer límites a la pobreza y a la desigualdad. Los trabajadores podrán demandar tal norma, pues la norma establecerá derechos económicos para todos —como, por ejemplo, el derecho al seguro del desempleo—. En el caso de una sociedad sigma, que es socialmente heterogénea, donde la ciudadanía está jerarquizada, existirán trabajadores que no tienen poder para establecer el contrato social del tipo que existe en épsilon. Ellos no tienen voz en el proceso político.

Comparado a épsilon, sigma será una sociedad con una mayor desigualdad que emerge del mercado y también con menos normas institucionales que pongan limitaciones a la desigualdad. En este contexto el comportamiento de los gobiernos de sigma será diferente al de épsilon: una mayor parte del presupuesto público será asignada a gastos discrecionales. La consecuencia es que el grado de desigualdad de equilibrio seguirá siendo mayor en sigma. Sigma mostrará, en suma, un mayor grado de desigualdad y un mayor grado de desorden social comparado a épsilon. Tenemos ahora otra proposición beta que se puede utilizar para refutar el modelo de la teoría de los gobiernos. Esta refutación la presentaremos en un modelo de equilibrio general más adelante. Pero antes necesitamos establecer el efecto del desorden social en el proceso productivo.

Desorden social y productividad laboral

La cantidad producida por las firmas depende, según el supuesto de la función de producción, de la cantidad de factores de producción que utilicen. Estos factores incluyen los que son propios de la firma —su stock de capital físico y tierra— y los que compran del mercado —mano de obra—. Introduciremos ahora un nuevo supuesto: los bienes públicos juegan un papel importante en el proceso productivo, a mayor cantidad de bienes públicos, mayor será el nivel del producto. Entre los factores productivos también se encuentran los bienes públicos.

Podemos distinguir tres tipos de bienes públicos que ingresan al proceso productivo de las firmas: capital de infraestructura (como las carreteras), orden social y capital humano. Ya dijimos que el orden social es un bien público. Puede parecer extraño que el capital humano sea un bien público cuando las firmas compran sus servicios en el mercado. Pero el capital humano es un bien público en el sentido que es producido colectivamente, vía la educación pública —financiada con impuestos—.

El supuesto de que a mayores cantidades de bienes públicos le corresponderá un mayor nivel de productividad promedio del trabajo —producto por trabajador— es clara con respecto a la infraestructura y al capital humano. La mayor cantidad de capital de infraestructura reduce los costos de producción de las firmas. Trabajadores con mayor capital humano utilizarán las máquinas con mayor eficiencia y aumentarán la productividad laboral.

El orden social como un factor de producción necesita mayor explicación. Dada las cantidades de capital y mano de obra de las firmas, la cantidad agregada producida dependerá del grado de orden social. Cuanto mayor el grado de desorden social, mayores serán las pérdidas de la producción de las firmas debido a las interrupciones en el proceso productivo —como huelgas—, así como por las redistribuciones privadas que harán los individuos que no toleran la desigualdad. Las firmas se verán sometidas a una especie de pago de impuestos, pero no al Estado, sino a personas privadas. Dada la cantidad de capital y trabajo utilizados, el producto total de la firma disminuirá; más importante aun, las ganancias disminuirán como resultado de la redistribución privada.

¿Cuál será la respuesta de las firmas? Como las firmas buscan maximizar la ganancia, tomarán acciones para reducir las pérdidas. Incorporarán al proceso productivo más bienes de capital y más trabajadores, así como también comprarán mayores cantidades de seguros, para proteger tanto la propiedad privada de la firma como el flujo de su producción. Estas acciones implican aumentar los costos fijos y por lo tanto los costos totales. Así, las firmas incurrirán en una magnitud de

sobrecostos que, sin embargo, tendrán el efecto de reducir las pérdidas netas que ocasiona la excesiva desigualdad. Y será entonces rentable incurrir en esos sobrecostos, pues la ganancia será mayor que si la firma no reaccionara.

La consecuencia de este comportamiento de las firmas es que utilizarán más capital y trabajo del que requieren por razones tecnológicas. La misma cantidad de producto será ahora producido con una mayor cantidad de capital y trabajo, innecesario desde el punto de vista de la tecnología. El producto por trabajador disminuirá.

En el agregado, este comportamiento individual de las firmas llevará a que el mismo ingreso nacional será producido con una mayor cantidad de capital y trabajo de la que es tecnológicamente necesaria. El producto por trabajador en el agregado se caerá. La economía pierde en eficiencia económica. La causa está en la excesiva desigualdad que generan los sobrecostos a las firmas. La excesiva desigualdad crea, a través del desorden social, un costo adicional al proceso productivo. Las sociedades más desiguales tenderán a tener un sistema productivo menos eficiente comparado al de las sociedades menos desiguales. El costo económico de la desigualdad en la sociedad es la pérdida de eficiencia del sistema productivo.

Equilibrio general con desorden social

En las teorías estudiadas en el capítulo anterior los modelos estáticos mostraban equilibrio general con orden social. En esos modelos se hizo el supuesto, aunque solo de manera implícita, que la desigualdad resultante del proceso económico era socialmente tolerada. Se suponía allí que los individuos que no toleraban esta desigualdad eran relativamente pocos, de modo que el desorden social era pequeño y se podía ignorar. Los precios y cantidades de equilibrio, el nivel de ingreso nacional y su distribución se repetían período tras periodo en tanto los valores de las variables exógenas se mantuvieran constantes.

Volvemos ahora a la pregunta inicial, ¿podría existir equilibrio general con desorden social? Consideremos el caso en que la distribución del ingreso nacional que resulta del proceso económico es tan desigual que genera un grado de desorden social. Los pobres intentarán restaurar la desigualdad hacia un grado más tolerable —ante el fracaso del gobierno de hacer una redistribución legal— y tomarán acciones para la redistribución privada —ilegal— del ingreso o de los activos. Estas reacciones implican violar algunas reglas del juego, como el respecto a los derechos de propiedad. Así se genera el desorden social.

La redistribución privada implicará transferencias netas de los ricos hacia los pobres en un monto global, será una transferencia forzosa y en un monto dado. Una especie de «impuesto» a los ricos, pero de carácter privado e ilegal. Este es un lado de la reacción a la excesiva desigualdad de un grupo de actores sociales y un componente de la posible solución de equilibrio general.

Pero las firmas también reaccionarán. Como se mostró anteriormente, las firmas buscarán la protección de sus propiedades utilizando más recursos de capital físico y trabajo del que es tecnológicamente necesario. Estas reacciones implican mayores costos fijos. En sus hogares, los ricos tomarán similares medidas. La industria de la seguridad se expandiría. Este es otro componente de la posible solución de equilibrio general.

Las reacciones del gobierno, que actúa guiado por su motivación de la maximización de votos y sujeto a la presión de los capitalistas, no reducirá la desigualdad primaria. Como se mostró arriba, la distribución primaria y secundaria no serán muy distintas. Los incentivos de los gobiernos van en la dirección de favorecer la represión antes que la redistribución, pues los efectos de la represión son más visibles e inmediatos. Así, el gasto público se reorientará a más gastos en represión —policías, jueces y cárceles— y menos a transferencias de ingresos o de activos que puedan reducir la excesiva desigualdad. Este es el tercer componente de la posible solución de equilibrio general.

Las acciones y reacciones provienen de los tres actores sociales: los trabajadores, los capitalistas y el gobierno. Ellos establecen con su comportamiento tres ecuaciones y deben resolver tres variables endógenas: el monto de la transferencia privada hacia los pobres, el monto de sobrecostos y el gasto público en represión. Podemos esperar que exista una solución a estas interacciones y que tengamos, entonces, un nuevo equilibrio general.

Hay que señalar que esas acciones y reacciones en el comportamiento de los tres actores sociales no afectarán, sin embargo, ninguna curva de demanda o de oferta en la economía. La razón es simple: los costos en que incurren las firmas son costos fijos, las transferencias privadas que logran los pobres son en montos globales, parecidos a un impuesto a la propiedad o al ingreso, y el gobierno solo reorienta el gasto público.

Por lo tanto, podemos suponer que los precios y cantidades de equilibrio que resultan del funcionamiento del sistema de mercado no se modificarán. Los precios y cantidades del equilibrio general se repetirán periodo tras periodo mientras las variables exógenas se mantengan fijas. ¿Cuál es entonces la diferencia con el caso en que había orden social? El equilibrio general con desorden social será con el mismo nivel del producto, pero producido con un mayor costo. Por lo tanto, el nivel de ganancias de las firmas será menor.

La otra diferencia está en la distribución del ingreso nacional. Denominamos distribución primaria a la que emerge del funcionamiento del mercado y distribución secundaria a la que resulta de las acciones del gobierno para modificar la primaria. También concluimos que en el equilibrio, ambas distribuciones no serían muy distintas, pues el gobierno no tiene poder ni incentivos para modificarla significativamente. Ahora debemos introducir una tercera categoría, la *distribución terciaria*. Esta resulta de la modificación que hacen los individuos a la distribución primaria por medio de la redistribución privada e ilegal.

La desigualdad será menor en la distribución terciaria que en la primaria, aunque la diferencia no será muy significativa. El ingreso de los trabajadores pobres aumentará debido a los ingresos no contractuales

que pueden obtener con la redistribución privada, *pero no tanto como* el que se necesitaría para reducir la distribución primaria y restaurar el orden social. Hay que notar que el ingreso que los individuos obtienen por medio de la redistribución privada está sujeto a riesgos, por ser precisamente no contractual o ilegal. Esto es debido a las mayores medidas de protección de la propiedad privada que aplican los capitalistas y a las medidas de represión del gobierno.

En el equilibrio general con desorden social, la distribución primaria se repite periodo tras periodo, pero sujeta a golpes redistributivos que sufrirán los ricos de manera aleatoria. Como consecuencia, en la distribución terciaria las ganancias disminuyen pero los ingresos laborales aumentan. En la economía de épsilon, el ingreso laboral aumenta porque el mayor empleo debido a esos golpes reduce el desempleo; en la economía de omega o sigma, el ingreso laboral aumenta debido a que el mayor empleo en el sector capitalista reduce el exceso de oferta y por lo tanto aumenta el ingreso promedio en el sector de subsistencia.

Cuando la distribución primaria es socialmente inaceptable, el equilibrio general de la economía será con desorden social. La reacción de los pobres hacia una redistribución privada hace que la economía funcione con golpes redistributivos aleatorios que sufrirán los ricos, represión del gobierno y sobrecostos de las firmas por la protección de la propiedad privada. Los gobiernos deben administrar este desorden social y son, por eso, vulnerables a la falta de legitimidad. La sociedad es políticamente inestable. Todos pierden en calidad de vida porque viven en una sociedad con desorden social, es decir, con violencia, ilegalidad, corrupción, burocracia y desconfianza.

Si por algún mecanismo se consiguiera reducir el excesivo grado de desigualdad que proviene del mercado, el desorden social disminuiría, y todos los actores ganarían. Todos vivirían en una sociedad con orden social, de mejor calidad. ¿Por qué entonces el grado de desigualdad no es el de tolerancia social, es decir, por qué el grado de desigualdad no se autorregula? El modelo muestra que ningún agente tiene el incentivo y el poder para hacerlo. Los trabajadores tienen el incentivo pero no

el poder. Los capitalistas tienen el poder pero no los incentivos, pues el orden social es un bien público y está entonces sujeto al problema olsoniano. El gobierno no tiene suficiente poder y no tiene el incentivo de modificar la distribución primaria. El equilibrio general es con desorden social, pero es, en efecto, una situación de equilibrio.

En el equilibrio general, dada las variables exógenas, las interacciones entre el gobierno, los capitalistas y los trabajadores llevarán a la sociedad a un equilibrio en el grado de desigualdad. El valor de este equilibrio puede estar fuera del rango de la tolerancia social y, por lo tanto, puede estar acompañado de desorden social. «Equilibrio» no implica aceptación de la situación por todos los actores sociales; implica, más bien, que nadie tiene ni el poder ni el incentivo para modificar la situación. El equilibrio distributivo tiene, así, la misma característica del equilibrio con desempleo que ocurre en el mercado laboral. El equilibrio distributivo implica que este valor se repetirá periodo tras periodo, siempre y cuando las variables exógenas se mantengan fijas.

Falsación de la teoría del desorden social

La predicción empírica, que se derivó de la teoría de la tolerancia limitada a la desigualdad, se refiere a que es que a mayor grado de desigualdad le corresponderá un mayor grado de desorden social. Esta proposición es, en principio, válida para todo tipo de sociedad capitalista.

La primera predicción empírica que proviene de los modelos de equilibrio general tiene que ver con diferencias entre las sociedades capitalistas. El cuarto capítulo mostró que la desigualdad en la distribución de ingresos —desigualdad en los flujos— depende positivamente de la desigualdad en la distribución de los activos —desigualdad en los stocks—. Dado que la desigualdad en la distribución de los activos es mayor en sigma que en épsilon y omega, se puede concluir que la desigualdad en la distribución de ingresos es mayor en sigma que en las otras sociedades.

Por lo tanto, la segunda predicción afirma que las diferencias en las desigualdades en activos económicos y sociales implican diferentes respuestas al desorden social. Así, en épsilon y omega, que son socialmente homogéneas, habrá incentivos para establecer normas que pongan límites a la desigualdad; en sigma, en cambio, no existirán incentivos para establecer tales normas por tratarse de una sociedad socialmente heterogénea y jerárquica.

La consecuencia es que la relación entre desigualdad y desorden social dependerá del tipo de sociedad. La relación es positiva, a mayor grado de desigualdad le corresponde un mayor grado de desorden social, pero los niveles serán distintos: una curva debajo de otra, la más alta para sigma y la más baja para épsilon, con omega en el medio. Amigo lector: usted puede construir su propio gráfico: mida en el eje horizontal el coeficiente de Gini, que va de cero a uno, y mida en el eje vertical el grado de desorden social, que también va de cero a uno; luego trace tres curvas de pendiente positiva, una debajo de la otra y señale el que corresponda a cada sociedad.

Si las sociedades tuvieran el mismo grado de desigualdad, habría mayor grado de desorden social en sigma. La razón es que en épsilon existen mecanismos institucionales para redistribuir el ingreso; en sigma, en cambio, no existe tal mecanismo y entonces los trabajadores solo pueden utilizar el desorden social como mecanismo para obtener alguna redistribución del ingreso. La democracia opera de manera distinta entre estas sociedades. El grado de democracia es entonces endógeno y depende de la distribución inicial de activos en la sociedad.

Dado que los grados de desigualdad son diferentes y que los mecanismos de redistribución son también distintos, la sociedad sigma operará en el segmento de alto grado de desigualdad y alto grado de desorden social, mientras que la sociedad épsilon operará en el rango de bajo grado de desigualdad y bajo grado de desorden social. La sociedad omega se ubicará al medio. Esta es la segunda predicción del modelo. Amigo lector: seleccione un punto en la parte baja de la curva de épsilon y otra en la parte superior de la curva sigma y únalos con otra curva,

la cual debe pasar también por el punto que representa a omega. Esta relación positiva es la predicción de la teoría.

Veamos ahora la refutación empírica. Las dos proposiciones beta establecidas arriba se pueden colocar en términos de las tres sociedades capitalistas bajo estudio. En primer lugar, la desigualdad en sigma es mayor que en épsilon. Si los países del primer mundo se definen como sociedades épsilon y los del tercer mundo como sigma —justificado por los resultados del capítulo anterior—, el grado de desigualdad es, en efecto, mayor en el tercer mundo. Esta predicción es consistente con la sétima regularidad del sistema capitalista. Los países del tercer mundo que se clasifican como omega son pocos y tienen una posición intermedia.

Queda por mostrar empíricamente que estas diferencias en la desigualdad en los flujos de ingresos están asociadas a diferencias en la desigualdad en la distribución de activos económicos y sociales entre el primer y el tercer mundo. En el segundo capítulo se argumentó que este es el caso.

En segundo lugar, la predicción de que, en general, el grado de desorden social de una sociedad será mayor cuanto mayor sea su grado de desigualdad implica que, en particular, el grado de desorden social será mayor en los países del tercer mundo que en los del primer mundo. Desorden social se mide, según el modelo teórico, por el grado de violación a los derechos de propiedad, el grado de inestabilidad política y el grado de democracia.

Los estudios empíricos internacionales sobre la violación de derechos de propiedad tienden a corroborar esta predicción. Un estudio que utiliza una muestra de 45 países —que incluyen el primer y tercer mundo— ha encontrado una relación estadística que indica que los países más desiguales tienen tasas de criminalidad mayores asociados a la propiedad; y en una muestra de 34 países, que los países más desiguales tienen tasas mayores de robos (Fajnzylber, Lederman y Loayza 2002). Otro estudio basado en una muestra de 50 países —del primer y tercer mundo— también encontró una relación estadística que señala

que países con mayores desigualdades tienen tasas de criminalidad mayores (Bourguignon 2000).

Sobre la desigualdad y la ilegalidad en los derechos de propiedad, no existen estudios empíricos específicos; sin embargo, la literatura internacional sí muestra que el tamaño de las actividades ilegales, el llamado «sector informal», en el tercer mundo es muy grande comparado al primer mundo. El sector ilegal produce bienes de consumo ilegal —tráfico de drogas, armas, trabajadores, contrabando de bienes—, pero también produce bienes que son de consumo legal y cuya demanda viene de las masas de bajos ingresos, es decir, como bienes inferiores. La desigualdad está subyacente en la existencia de la ilegalidad y constituye un medio de redistribución privada del ingreso.

Sobre la relación entre desigualdad e inestabilidad política, también existe evidencia empírica consistente con la predicción del modelo. En efecto, un estudio empírico ha encontrado una relación estadística positiva entre estas variables en una muestra de setenta países del primer y tercer mundo con datos del período 1960-1985 (Asesina y Perotti 1996). Y otro estudio encontró una relación estadística negativa entre el grado de democracia y el grado de desigualdad, utilizando datos de una muestra de 55 países del primer y tercer mundo para los años 1960 y 1980 (Muller 1997).

La literatura internacional también muestra la debilidad de las instituciones en el tercer mundo. Las normas no se cumplen, el poder judicial es débil, la corrupción es amplia. Todas estas anomalías del sistema capitalista también se explican por el modelo teórico desarrollado aquí. El desorden social implica debilidad de las instituciones, que se expresan en violencia, ilegalidad, corrupción y burocratismo. La diferencia es que en esa literatura se considera que la debilidad de las instituciones es exógena, es causa de los males que agobian al tercer mundo. En este estudio, en cambio, esa debilidad es endógena. La debilidad de las instituciones no es causa sino consecuencia de los males del tercer mundo, entre los que se incluye la excesiva desigualdad en los activos económicos y políticos que dan lugar a una excesiva desigualdad en los ingresos.

Estos resultados empíricos muestran que, en general, tanto el grado de desigualdad como el grado de desorden social son mayores en el tercer mundo en comparación al primer mundo. Los países del primer mundo han desarrollado sistemas de protección social que pone límites a la desigualdad. Derechos económicos como el seguro al desempleo constituyen un claro ejemplo. En el tercer mundo casi no existen tales derechos y, por lo tanto, el exceso de desigualdad es mucho más probable y conduce al desorden social.

Conclusiones

La introducción de la teoría de la tolerancia limitada a la desigualdad, así como la de la teoría de los gobiernos en el equilibrio general, nos ha permitido construir un modelo de equilibrio general con desorden social. Hemos mostrado que una sociedad con excesiva desigualdad tendrá un equilibrio general, pero será con desorden social. Será una sociedad con más baja calidad de vida y más baja productividad laboral comparada con una sociedad sin excesiva desigualdad.

El funcionamiento del mercado determinará los precios y cantidades de equilibrio, así como la distribución primaria del ingreso. Este será el equilibrio general del mercado. Los precios y cantidades prevalecerán, pero no así la distribución primaria del ingreso. Habrá una redistribución del ingreso, o de los activos, producto de la reacción de la gente al intolerable grado de desigualdad, quienes buscarán entonces redistribuir la distribución primaria por medios privados y forzosos. Pero la redistribución no afectará ni los precios ni las cantidades del equilibrio general inicial. Por lo tanto, ese equilibrio se repetirá periodo tras periodo en tanto las variables exógenas se mantengan fijas, pero sujeto a los golpes redistributivos privados.

Las sociedades pueden entonces funcionar con distintos grados de orden social. Y esto es precisamente lo que observamos en el mundo real. ¿Cómo explica la teoría de desorden social estas diferencias? El

grado de desorden en una sociedad depende de su grado de desigualdad en la distribución de ingreso y de las normas institucionales sobre la redistribución de ingresos. También dice que ambos factores dependen de la desigualdad inicial en la distribución de los activos. Por lo tanto, la teoría predice que las sociedades que tienen una mayor desigualdad en la distribución inicial de los activos económicos y sociales funcionarán con un mayor grado de desorden social.

La sociedad sigma tiene un mayor grado de desigualdad inicial que las otras y como los países del tercer mundo se parecen a esta sociedad —como se mostró en el capítulo anterior—, el modelo predice que este grupo de países funcionará con mayor grado de desigualdad y con mayor grado de desorden social que los del primer mundo, que se parecen a la sociedad épsilon. La evidencia empírica existente no refuta estas predicciones del modelo de la teoría del desorden social. Por lo tanto, no hay razón alguna para rechazar esta teoría en esta etapa de nuestra investigación.

Así se puede explicar el por qué de las diferencias que observamos en el funcionamiento de las sociedades del primer y tercer mundo. La desigualdad inicial no puede ser vista como una cuestión ética solamente; también juega un papel central en la provisión de un bien público: el orden social, que es un determinante de la calidad de la sociedad en la que uno vive.

¿Tendrá la desigualdad inicial también un efecto sobre el desarrollo económico de la sociedad capitalista? En particular, ¿cuál es su papel en la acumulación de capital físico en las sociedades capitalistas? A responder esa pregunta se dedica el siguiente capítulo.

Capítulo 6

Acumulación del capital físico

En un modelo estático, el stock de capital físico de la sociedad aumenta solo de manera exógena. Bajo este tipo de modelo el nivel del producto crece de una sola vez y allí se mantiene hasta que el próximo aumento en el capital ocurra. Pero, un modelo estático no puede explicar el cambio continuo en el nivel de producción de las sociedades, es decir, no puede explicar el crecimiento económico como proceso. Los modelos presentados en los capítulos anteriores han sido todos estáticos.

Para explicar el crecimiento económico se necesita un modelo dinámico, donde el stock de capital aumente continuamente, de modo que el stock de un periodo determina el stock del siguiente y así sucesivamente; de esta manera, el stock de capital de la sociedad deviene en variable endógena.

La acumulación del capital físico se lleva a cabo mediante el gasto que hacen los capitalistas en la compra de bienes de capital. A este gasto se le denomina *inversión*. Los capitalistas compran bienes de capital por dos razones: primero, para reponer el capital desgastado —que se llama la depreciación— y así mantener la capacidad productiva de la firma constante; segundo, para aumentar el stock de capital y así elevar la capacidad productiva de la firma —que se llama inversión neta—. La inversión total se puede descomponer entonces en inversión de reposición y en inversión neta.

En el equilibrio general estático, la inversión de reposición o depreciación que hacen las firmas está incluida en el proceso productivo

como un costo fijo. El producto total o ingreso nacional es, por lo tanto, igual al producto total neto, deducida la cantidad dedicada a la reposición del stock de capital desgastado. Esto es lo que se hizo en los modelos presentados en los capítulos anteriores, aunque solo de manera implícita. Por eso, podíamos hablar de proceso de producción, y la cantidad producida de equilibrio se podía repetir periodo tras periodo. En este capítulo y en los siguientes la depreciación estará deducida de la producción de manera explícita.

El objeto de estudio ahora será la inversión neta, que por comodidad llamaremos simplemente inversión. Este estudio se referirá al comportamiento de los capitalistas en su papel de inversionistas, de agentes de la acumulación de capital físico y, por lo tanto, de agentes del crecimiento económico de la sociedad. La teoría que se presentará a continuación intenta ser una teoría general, válida para las tres sociedades capitalistas: épsilon, omega y sigma.

Racionalidad de los inversionistas

En cuanto a la inversión, el capitalista debe decidir sobre varias cosas, tales como: ¿Cuánto invertir?, ¿en qué invertir?, ¿cómo financiar la inversión?, ¿en dónde invertir, es decir, en qué países? El interés central de este capítulo está en responder la pregunta de dónde invertir. ¿Cuánto invertir en los países del primer mundo y cuánto en los del tercer mundo? Se presentará entonces una teoría que busca explicar la inversión relativa —su asignación por países— y no el nivel absoluto de la inversión.

La teoría estándar supone que los inversionistas buscan la máxima tasa de retorno de sus inversiones. Debido a los rendimientos decrecientes del factor capital, países con poco capital tendrán tasas de retorno más altas que países con abundante capital. Luego, una primera predicción que se puede establecer es que la tasa de retorno de una inversión en un país depende de sus dotaciones de capital físico. Como

los países del tercer mundo son los menos dotados de capital, los retornos de la inversión deben ser mayores que en el primer mundo; como consecuencia, la inversión mundial debe dirigirse principalmente al tercer mundo.

Esta predicción es refutada por los datos: la inversión mundial se dirige en su mayor parte al primer mundo, no a los países del tercer mundo. En la última década, apenas el 20% de la inversión extranjera se dirigió al tercer mundo (Markussen 2002; UNCTAD 2006).

Robert Lucas, profesor de economía de la Universidad de Chicago y ganador del Premio Nobel, ha tratado de salvar la teoría estándar introduciendo otro factor en el retorno de la inversión: la dotación de capital humano entre países. Los países del primer mundo están dotados no solo de más capital físico, sino también de más capital humano. La productividad del capital físico aumenta cuando la operan trabajadores con mayor capital humano y, por lo tanto, el retorno de la inversión en capital físico debe ser mayor allí donde exista una mayor dotación de capital humano, es decir, en el primer mundo. Así se explicaría por qué la inversión mundial se dirige allí. Ahora, los datos observados no refutan sino, por el contrario, corroboran esta predicción de la teoría estándar.

Existe un problema lógico en el argumento del profesor Lucas. Suponiendo, como él lo hace, que el nivel de la producción de un proyecto depende de la cantidad de capital físico y del capital humano, y de la cantidad apropiada de trabajadores, la productividad marginal del capital dependerá del stock de capital humano y del capital físico que exista en la economía. El primer efecto haría que la productividad sea mayor en el primer mundo, pero el segundo factor haría que la productividad sea mayor en el tercer mundo. Luego, los dos efectos tienden a compensarse y la conclusión del profesor Lucas no es una necesidad lógica. No es una hipótesis refutable.

Queda, entonces, la pregunta, ¿cuáles son los determinantes del comportamiento de los inversionistas en cuanto a elegir inversiones entre países? Aquí desarrollaremos una teoría del comportamiento de los inversionistas. Procederemos al proceso de falsación más adelante.

La teoría que desarrollaremos se refiere a entender la decisión de los inversionistas entre invertir en una sociedad sigma, omega o épsilon. Tal teoría de inversión supondrá que el capital financiero es perfectamente movible en la economía mundial; es decir, no existen restricciones o prohibiciones para que los inversionistas puedan invertir donde quisieran hacerlo.

También supondremos que los inversionistas toman sus decisiones en un contexto de incertidumbre. Los rendimientos de las inversiones no son seguros, y así como pueden obtener ganancias, también pueden sufrir pérdidas. Las inversionistas corren entonces riesgos, esto es, el riesgo de sufrir pérdidas. Supondremos que los inversionistas conocen los distintos eventos que pueden ocurrir y los rendimientos en cada evento, así como las probabilidades de que los eventos ocurran. Toda esta información se resume en dos números que se supone ellos conocen: el rendimiento esperado y el grado de riesgo —medido por la desviación estándar de los retornos— de un gasto en inversión.

El supuesto sobre las motivaciones detrás del comportamiento de los inversionistas es que persiguen el máximo rendimiento esperado y el mínimo riesgo de sus inversiones. Preferirán aquellos proyectos de inversión que, a igualdad de condiciones en el grado de riesgo, tienen mayor rendimiento esperado; de manera similar, a igualdad de rendimientos esperados, preferirán aquellos que ofrezcan un menor grado de riesgo.

En casos donde unos proyectos no muestran superioridad sobre otros, donde aquellos que tiene mayor rendimiento esperado son más riesgosos comparado a los otros, el comportamiento del inversionista dependerá de sus preferencias. El supuesto de la teoría estándar es que los inversionistas tienen aversión al riesgo; es decir, ellos están dispuestos a sacrificar algo de rendimiento esperado con tal de tener un nivel menor de riesgo. Bajo este supuesto, ellos no invertirán en un solo proyecto sino en varios. Así la cartera o portafolio del inversionista estará diversificado y tendrá, entonces, un nivel de riesgo intermedio, aunque también con un rendimiento esperado que es intermedio. La diversificación será la condición de equilibrio en el comportamiento de los inversionistas.

Establecimos en el cuarto capítulo el supuesto de que la racionalidad de los capitalistas incluye dos objetivos: mantener su posición en la clase capitalista y obtener la máxima ganancia, donde el primer objetivo tiene prioridad. Consistente con esta racionalidad, los capitalistas buscan evitar pérdidas económicas que les hagan perder su membresía en la clase capitalista.

Si un capitalista tiene una riqueza de cien millones de dólares, no puede entrar en juegos —como hacer inversiones— donde pudiera perder más de esa cantidad, pues si ello ocurriera dejaría de ser capitalista. Si para operar como capitalista de su clase necesita tener ochenta millones de dólares, entonces no puede entrar en juegos donde pudiera perder más de veinte millones. Los capitalistas enfrentan entonces un riesgo que es soportable, cuando las pérdidas, de ocurrir, podrían ser absorbidas por la riqueza que poseen; y otro insoportable, cuando las pérdidas no podrían ser absorbidas, que llevarían al desastre económico y social, que consiste en dejar de ser capitalista. El capitalista evitaría entrar en juegos que lo puedan llevar al desastre económico.

En la teoría estándar, los inversionistas no tienen límites a los riesgos de sufrir pérdidas. Por contraste, aquí supondremos que la racionalidad del inversionista es más sofisticada: toma decisiones de manera secuencial. Primero, como el inversionista tiene aversión a los juegos riesgosos, que le pueden llevar al desastre económico y social, evita los proyectos que tengan riesgos insoportables, independiente de sus retornos esperados; además, su capacidad de absorber pérdidas depende de su nivel de riqueza. Segundo, entre los proyectos que tienen riesgos soportables, elige un portafolio de acuerdo a sus preferencias. El capitalista, en suma, no invertirá en proyectos donde el riesgo de pérdidas sea insoportable. Así asegura su posición de clase. Fuera de esos proyectos, el capitalista invertirá en los proyectos que le den la mejor combinación de rendimiento esperado y riesgo (soportable) que a él más le guste.

Esta teoría del inversionista supone que la capacidad de absorber pérdidas, de asumir riesgos, depende de la riqueza del inversionista. Luego, la teoría predice que el comportamiento de los inversionistas

será diferente, dependiendo del tamaño de su riqueza. Los grandes inversionistas tendrán un portafolio de proyectos distintos, con mayores rendimientos esperados y riesgos, comparado al de los pequeños.

¿Dónde invertir: en el primer o en el tercer mundo?

Con esta racionalidad de los inversionistas, tratemos ahora de explicar la regularidad empírica anotada arriba: la baja proporción de inversiones extranjeras que se dirige hacia el tercer mundo. Para tal efecto, construiremos la teoría de la inversión relativa —sobre las proporciones relativas que se dirigen a cada tipo de economía—. Esta teoría toma como dada, como variable exógena, el nivel de la inversión total.

El supuesto primario de la teoría es que los inversionistas determinan su portafolio de inversiones en los distintos tipos de economías guiados por la racionalidad mencionada anteriormente. Para refutar esta teoría necesitamos construir un modelo. Seguiremos suponiendo que el sistema capitalista se compone de las tres sociedades capitalistas abstractas que hemos estudiado, pero introduciremos un nuevo supuesto: los rendimientos esperados dependen negativamente de la dotación de capital por trabajador de las sociedades y positivamente de los bienes públicos, tales como capital humano, infraestructura y orden social; mientras que el grado de riesgo depende negativamente del orden social. Debido a que los rendimientos esperados y el grado de riesgo no son observables; ahora introducimos supuestos sobre los factores observables que los determinan.

Dada su racionalidad, los inversionistas decidirán la asignación de sus fondos de inversión a las tres sociedades de acuerdo a estos factores determinantes del rendimiento esperado y del riesgo, que ahora constituyen variables exógenas del modelo. La otra variable exógena será el nivel de la inversión mundial total. Por pura comodidad, consideremos solo épsilon y sigma por ahora. Consideremos que en cada sociedad existe un conjunto de proyectos de inversión con su rendimiento

esperado y su grado de riesgo, que son conocidos por los inversionistas. Si consideráramos solo al grupo de capitalistas para quienes estos riesgos son soportables, podemos distinguir tres casos posibles:

Si los proyectos en épsilon tuvieran el rendimiento esperado más alto y el riesgo más bajo comparado a los proyectos en sigma, los inversionistas colocarían todos sus fondos en épsilon, pues sigma no podría competir con épsilon.

Si los proyectos en sigma tuvieran un mayor rendimiento esperado alto y un grado de riesgo más bajo, los inversionistas colocarían todos sus fondos en sigma.

Si los proyectos en épsilon tienen un rendimiento esperado bajo pero un riesgo bajo, mientras que los proyectos en sigma tienen un rendimiento esperado alto pero también un riesgo alto, entonces los proyectos en épsilon y sigma compiten en la cartera de los inversionistas. Una parte de los fondos de inversión se irán a proyectos en épsilon y otra a proyectos en sigma; es decir, habrá diversificación.

Los dos primeros casos son refutados por los datos. El tercero es el que muestra consistencia con la regularidad empírica mencionada anteriormente. ¿Cómo se explica este resultado?, ¿cómo se explica la composición de la cartera de los inversionistas, donde la mayor proporción se va a épsilon?

Recordemos que las dotaciones factoriales son tales que épsilon está más dotado de capital físico por trabajador que sigma. Luego, por efecto de esta variable, el rendimiento esperado de la inversión en capital físico será menor en épsilon que en sigma, debido al efecto de los rendimientos decrecientes. Pero épsilon también está dotada de más bienes públicos —infraestructura, capital humano, orden social— que sigma, lo cual hará que la curva de productividad del capital se eleve y el rendimiento esperado sea mayor en épsilon. Luego, el rendimiento esperado de la inversión en capital físico en épsilon tenderá a ser ambiguo con respecto al valor en sigma, pues el efecto de los rendimientos decrecientes y el de los bienes públicos tienden a compensarse. Como el orden social es mayor en épsilon que en sigma, el riesgo será mayor

en sigma que en épsilon. En suma, la diversificación en sociedades, se puede derivar de los supuestos del modelo, pero no es una necesidad; el modelo, por tanto, no es refutable.

Si añadimos el supuesto de que algunos proyectos son específicos a una de las sociedades, como sería el caso de proyectos de inversión en explotaciones mineras y de otros recursos naturales en los países del tercer mundo, el rendimiento esperado será mayor en el conjunto de proyectos en las sociedades sigma. El modelo ahora sí conduce a una situación de equilibrio con diversificación.

Según el modelo, la inversión mundial se destinará a épsilon y a sigma, pues los proyectos de ambas economías compiten en la cartera de los inversionistas. Pero los inversionistas enfrentarán precios relativos poco favorables para invertir en sigma. La diferencia en rendimientos esperados es pequeña —debido a efectos que tienden a compensarse— confrontada a la gran diferencia en los niveles de riesgos entre épsilon y sigma. Luego, obtener una unidad mayor de rendimiento esperado será muy costoso en términos del riesgo que deben de asumir. El bien rendimiento esperado es un bien muy costoso en unidades del bien seguridad y por lo tanto los inversionistas no desearán comprar mucho de ese bien, lo cual implica invertir poco en sigma. Así, la teoría predice una baja proporción de la inversión mundial que va a sigma. Esta es la condición de equilibrio, dado los valores de las variables exógenas.

Predicciones empíricas y falsación

Los cambios en las variables exógenas tendrán el efecto de modificar la cartera de los inversionistas, es decir, las proporciones de las inversiones que van a cada sociedad en cada periodo. Nótese que la composición de la cartera del inversionista se refiere a la distribución del stock de su riqueza entre épsilon y sigma, mientras que los flujos de inversiones en cada sociedad significan cambiar la composición de esa cartera o portafolio.

Debido a que la decisión que toman los inversionistas sobre dónde dirigir sus inversiones toma en cuenta las situaciones en ambas sociedades, esta decisión es simultánea. Luego, para entender la composición de la cartera del inversionista, es suficiente conocer los determinantes de la proporción que va a sigma, pues la proporción restante irá a épsilon.

El modelo predice que un aumento del stock de capital humano en sigma relativo a épsilon elevará la proporción de sigma en el portafolio de los inversionistas. Mayor cantidad de capital humano en la economía aumenta la productividad del capital físico y, por lo tanto, en el rendimiento esperado. Luego, la proporción de sigma aumentará. Igual relación predice con respecto a la infraestructura. El modelo también predice similar efecto con respecto al grado de orden social. El mayor grado de orden social no solo eleva la productividad del capital y el rendimiento esperado sino también reduce el riesgo. ¿Cómo se puede aumentar el grado de orden social? De acuerdo a la teoría del desorden social, reduciendo el grado de desigualdad en la sociedad.

El nivel de los fondos de inversión mundial es exógeno en este modelo. Los capitalistas pueden recurrir a distintas fuentes de financiamiento para llevar a cabo sus inversiones en capital físico, tales como las ganancias de sus propias firmas, los mercados de crédito y los mercados de capitales financieros —mediante la emisión de bonos o de acciones—. Los cambios en esta oferta de fondos afectarían el nivel de la inversión mundial, pero mantendremos el supuesto de que el nivel de la inversión mundial es exógena.

La última predicción se refiere al efecto de un aumento en el nivel de riqueza de los capitalistas sobre el portafolio agregado entre épsilon y sigma. Este efecto dependerá no solo del nuevo nivel de riqueza sino también de su composición. Si la riqueza de los grandes capitalistas aumenta en mayor proporción que la de los pequeños, entonces las inversiones más riesgosas aumentarán, pues el nivel de riqueza de los inversionistas es mayor y pueden entonces invertir en proyectos más riesgosos que antes no podían, ya que ahora esos mayores riesgos de pérdidas son absorbibles o soportables. Como resultado, las inversiones

en sigma aumentarán en mayor proporción que en épsilon. Si la riqueza de los capitalistas pequeños es la que aumenta, el efecto irá en la dirección contraria. Si la composición de la riqueza de los capitalistas se mantiene, el efecto sobre el portafolio tenderá a ser neutral.

Hasta ahora hemos ignorado la existencia de la sociedad omega. Dado que sus dotaciones factoriales se ubican entre épsilon y sigma; y dado que la desigualdad inicial es más parecida a épsilon, la teoría predice que omega ocupa en la cartera de los inversionistas una proporción más cercana a épsilon que a sigma. Pero, según los resultados del cuarto capítulo, la mayoría de los países del tercer mundo se parecen a la sociedad abstracta sigma.

Existen estudios empíricos en la literatura internacional que nos permite hacer la confrontación empírica de las predicciones del modelo. Sobre la predicción de que el *nivel* de la inversión, que se dirige a los países depende positivamente de su dotación de capital humano y de su grado de igualdad en la distribución del ingreso nacional, la literatura internacional muestra resultados consistentes. Así, el estudio de Barro y Sala-i-Martin (1995) encontró una fuerte correlación positiva entre inversión privada, educación —una medida de capital humano— y estabilidad política —asociada a la desigualdad inicial en nuestra teoría— en una muestra de 134 países para el periodo 1965-1985. Markusen (2002) encontró una correlación positiva entre inversión privada y educación por países en el comportamiento de las firmas multinacionales que operan en el primer y tercer mundo a la vez. Alesina y Perotti (1996) encontraron una correlación positiva entre inversión privada y estabilidad política, así como una correlación negativa entre estabilidad política y grado de desigualdad en la distribución del ingreso nacional (medido para 1960), en una muestra de setenta países del primer y del tercer mundo, para el periodo 1960-1985.

La predicción de que la *proporción* de la inversión que va al tercer mundo es muy baja comparada a la del primer mundo, y que se debe a las diferencias en las dotaciones factoriales y en el grado de la desigualdad inicial es consistente con la evidencia empírica mostrada al inicio.

Por lo tanto, los aumentos observados en el *nivel* de inversión en el primer y en el tercer mundo se debieron, sobre todo, al aumento en el nivel mundial de inversión, pues las diferencias en las dotaciones factoriales y en el grado de desigualdad no han cambiado mucho. El modelo de la teoría de la inversión relativa desarrollada aquí no es refutado por los datos de la realidad. Por lo tanto, no hay razón alguna para rechazar la teoría de la inversión relativa en esta etapa de nuestra investigación.

OTRAS PREDICCIONES DEL MODELO

El resultado de que la asignación de la inversión privada depende positivamente del orden social, que a su vez depende de la desigualdad inicial, nos lleva a concluir que la inversión privada en los países depende negativamente de su desigualdad inicial. La teoría de la inversión relativa predice esta conexión y esta predicción está corroborada por la evidencia empírica existente.

La relación negativa entre inversión privada y desigualdad inicial nos permite explicar varios hechos recurrentes de la realidad internacional. Algunos de ellos los presentamos aquí.

Primero, está la cuestión de la competencia entre los países por ganar los mercados internacionales de bienes. En la economía estándar se utiliza la teoría de las ventajas comparativas para responder esta pregunta. Según esta teoría del comercio internacional, la competitividad de los países en los mercados internacionales de bienes depende de sus productividades relativas. Pero en la misma teoría estándar existe otra teoría macroeconómica que sostiene que, dada las productividades relativas, la competitividad dependerá del tipo de cambio, pues las variables nominales tienen efectos sobre las variables reales. Aunque estas teorías no aparecen como una teoría unitaria en la economía estándar, se podría decir que la primera teoría se refiere al largo plazo y la segunda al corto plazo.

Pero, ¿es la competitividad de largo plazo exógena o endógena? La mayor inversión en capital físico en una industria aumenta la productividad laboral, la cual incrementa el grado de competitividad del país en esa industria en el comercio internacional. Pero según la teoría de la inversión relativa presentada aquí, la inversión es endógena y depende, entre otros factores, del grado de desigualdad de los países: países con un menor grado de desigualdad obtendrán una mayor proporción de la inversión mundial, y por esta vía tendrán una mayor competitividad en un mayor número de industrias. En suma, la competitividad es endógena y los países compiten entre sí en los mercados internacionales de bienes con sus grados de igualdad.

La mayor participación de los países llamados «tigres asiáticos» en el comercio internacional es consistente con esta predicción. Recordemos que la mayoría de los países asiáticos tienen un grado de desigualdad que es menor al de los demás países del tercer mundo. Algunos son países tipo omega, como los clasificamos en el cuarto capítulo.

Segundo, está la cuestión de la estructura productiva de los países. La teoría estándar del comercio internacional predice que países abundantes en mano de obra se especializarán en la producción de bienes intensivos en mano de obra. El comercio de bienes es un sustituto perfecto de la movilidad de factores, y así los salarios reales se igualarán entre países. La inversión privada se dirigirá entonces a proyectos intensivos en mano de obra y así el exceso de oferta laboral se reducirá. La evidencia empírica refuta estas predicciones. La sobrepoblación no está reduciéndose, a pesar de que las tasas de crecimiento poblacional han disminuido casi en todo el tercer mundo. La falla puede deberse a que la teoría estándar ignora el papel de la desigualdad tanto en el funcionamiento de la economía como en la inversión.

Dado que el orden social es un factor de producción, se puede distinguir entre industrias que son más intensivas o menos intensivas en este factor, de la misma manera como se distingue entre industrias intensivas en capital o mano de obra. Industrias intensivas en orden social serán aquellas que tienen interacciones importantes con otras industrias

de la economía, tales como las manufacturas. Industrias menos intensivas en orden social, en cambio, serán aquellas que pueden ser producidas en enclaves, tales como centros mineros, centros petrolíferos, centros turísticos. La sociedad puede estar pasando por un importante desorden social y, sin embargo, estas industrias seguirán con su actividad productiva sin mayores perturbaciones.

Por otro lado, la inversión en proyectos de gran escala en los países del tercer mundo implica para los inversionistas asumir grandes riesgos. Estos proyectos solo podrán ser tomados por capitalistas con tamaños grandes de riqueza, que puedan absorber grandes pérdidas sin tener que llegar al desastre económico y social. Pero estos grandes proyectos se harán menos riesgosos, y más atractivos, si se van a producir en enclaves.

La otra predicción de la teoría de la inversión relativa es que en los países del tercer mundo la mayor parte de la inversión extranjera será de grandes firmas y producirán bienes relativamente menos intensivos en orden social. Esta predicción es consistente con un hecho que observamos en la economía capitalista mundial: la inversión extranjera en el tercer mundo se dirige a explotar los recursos naturales —minería, petróleo, gas— que tienen esas características. Esta inversión no se dirige a crear la industrialización de los países del tercer mundo, excepto por la producción en maquilas, que en realidad buscan explotar la mano de obra barata como recurso natural.

Debido a su efecto sobre la inversión, el grado de desigualdad de los países es un factor que afecta su estructura productiva. Dado el alto grado de desigualdad de los países del tercer mundo, será muy difícil que la inversión pueda cambiar su estructura productiva y su lugar en la división internacional del trabajo; será, además, muy difícil que puedan cambiar su grado de desigualdad inicial. La inversión extranjera generará cambios cuantitativos importantes en el tercer mundo, pero no podrá generar cambios cualitativos. El crecimiento económico no podrá modificar la desigualdad. Esta predicción será retomada en el capítulo sobre desarrollo económico.

La teoría de la inversión relativa predice, en suma, que los países compiten en dos esferas en la economía internacional: en el mercado de bienes y en el mercado de inversiones. La competencia en la segunda afecta la primera, pero no al revés. En la competencia por atraer inversiones, los países compiten con sus dotaciones iniciales, que incluyen la dotación inicial de factores productivos y la desigualdad inicial.

Mercado del crédito

Habíamos mostrado en el cuarto capítulo, donde utilizamos modelos estáticos, que el equilibrio general en las sociedades capitalistas era con desigualdad. Mostramos allí que esta situación era de equilibrio, pues no había actor social que tuviera el poder y el deseo de cambiarlo. En esos modelos no había mercado de crédito ni de seguros. ¿Cambiaría la solución si introdujéramos en el modelo estos mercados?

La idea es que en una sociedad tipo épsilon, donde todos están dotados del mismo capital humano, tanto capitalistas como trabajadores, no debería existir desigualdades en el largo plazo. Los capitalistas obtienen el mismo ingreso que los trabajadores por su capital humano. La diferencia en ingresos se debe a que ellos reciben, en adición, las ganancias que vienen por la propiedad de su capital físico. Si los trabajadores pudieran obtener crédito, podrían acumular capital físico y así devenir en capitalistas. ¿Qué hace, entonces, que las clases sociales se reproduzcan?

Necesitamos entender el funcionamiento del mercado de crédito. En este mercado se intercambia el bien crédito, donde el precio de mercado es la tasa de interés nominal. Construiremos aquí un modelo del mercado de crédito, donde supondremos que la estructura del mercado es de competencia perfecta. El mercado de crédito tiene por el lado de la demanda a los inversionistas que desean obtener préstamos para sus proyectos de inversión, y por el lado de la oferta a los bancos.

Por el lado de la demanda de crédito, supondremos que existe una curva de demanda decreciente: a menor tasa de interés, los inversionistas

desearán adquirir cantidades mayores de crédito para financiar sus proyectos. La razón es simple: los proyectos tienen diferentes tasas de retorno esperadas; además, la cantidad de proyectos de inversión que tienen una tasa de retorno de 5%, cuando menos, será mayor que la de 15%, puesto que este grupo de proyectos incluye al primero. Se supone que los proyectos son rentables y que harían posible el repago del préstamo.

Necesitamos ahora una teoría sobre la racionalidad de los bancos. Los bancos son empresas intermediarias, que toman dinero prestado de un grupo social y se lo prestan a otro grupo social. Su objetivo último, por supuesto, no es hacer la intermediación entre los que tienen dinero en exceso y los que son deficitarios; su objetivo es maximizar la ganancia en esta actividad. Luego, la tasa de interés activa —que cobran a los prestatarios— debe ser mayor que la tasa de interés pasiva —que pagan a los depositantes—; además, esta diferencia debe ser suficientemente alta como para cubrir los costos de la intermediación y así obtener ganancias.

Pero la intermediación es un negocio riesgoso. En particular, existe el riesgo de que los prestatarios no devuelvan el crédito. Los bancos no conocen bien a los que solicitan crédito; en realidad, el banco no sabe tanto sobre el prestamista y sus proyectos como lo sabe el propio prestamista. Este es el *problema de la información asimétrica.*

Los bancos tendrían que utilizar un mecanismo para protegerse del riesgo de no repago del préstamo. Un mecanismo es solicitar al prestatario una garantía, llamado el colateral, para que este tenga el incentivo de devolver el préstamo, pues si no lo hace el costo será perder la garantía. Además del colateral, los bancos toman un conjunto de acciones para reducir el problema de la información asimétrica y conocer mejor a sus clientes y a sus proyectos, acciones que involucran costos. Estos son los costos de transacción y se refieren a actividades tales como la información sobre el cliente mismo, sobre su historia crediticia, sobre el proyecto y también sobre la garantía y la ejecución de la garantía.

¿Pero cuáles son los objetos que puede presentar el prestamista como colateral? Naturalmente, su capital físico. En consecuencia, el

que no tiene una dotación de capital físico, no tiene nada para entregar como colateral y no puede acceder al crédito bancario, por más que tenga un proyecto rentable. Como no están dotados con capital, los trabajadores quedarán excluidos del mercado de crédito. Luego, para acumular capital y llegar a ser capitalista, ¡primero hay que tener una dotación de capital!

El otro mecanismo que utilizan los bancos es la exclusión: no todo el que tiene capital físico puede acceder al crédito. Supongamos que el colateral que piden los bancos sea igual al valor de préstamo. ¿Por qué entonces podría suceder que no todos los que poseen capital obtienen el crédito que desean transar? Supongamos que los costos de transacción no dependen del monto de crédito. El costo en que incurre el banco para evaluar y administrar el crédito será el mismo para una solicitud que pide mil dólares o cien mil dólares. Sin embargo, el *costo por dólar* será muy diferente: será mucho más costoso para el banco otorgar un crédito de mil que un crédito de cien mil. Como el banco busca la máxima ganancia, preferirá otorgar un único crédito de cien mil que cien créditos de mil. El banco preferirá dar pocos créditos grandes en lugar de muchos créditos pequeños. Los bancos no tienen incentivos para operar en un mercado de crédito al por menor, como un supermercado.

El banco evitará dar préstamos pequeños y pondrá un umbral al tamaño del crédito, que es equivalente a poner un umbral de capital físico para ser elegible como prestamista. Entonces los individuos dotados de capital físico que sea inferior al del umbral quedarán también excluidos del mercado de crédito, al igual que aquellos que no tienen capital físico. Esta es la teoría sobre la racionalidad de los bancos. Ahora veamos el funcionamiento del mercado de crédito.

La demanda de crédito tendrá dos componentes. La *demanda total* es la que proviene de todos los individuos que tienen un proyecto rentable y que desean intercambiar en el mercado de crédito. Pero luego está el componente de la *demanda efectiva*, que es la que proviene de los clientes elegibles solamente, es decir de aquellos que tienen una

dotación de capital físico por encima del valor de umbral. Los clientes elegibles para obtener créditos solo son los del segundo grupo. Luego, para acumular capital físico y llegar a ser capitalista, ¡primero hay que tener una dotación importante de capital!

Las interacciones entre el mercado de crédito y de depósitos establecerán los dos precios —tasa de interés pasiva y activa— y las dos cantidades —depósitos y créditos—. En el mercado de crédito, la curva de demanda de crédito es decreciente: a menor tasa de interés activa, mayor será la cantidad demandada de crédito, pues los proyectos que antes no eran rentables, ahora lo serán; y la curva de oferta será creciente, pues los costos variables de los bancos aumentará con la cantidad de créditos —con cantidades por prestamista superiores al umbral, por supuesto—.

En el mercado de depósitos, la curva de oferta proviene de los depositantes y es creciente: a mayor tasa de interés pasiva, mayor será la cantidad de crédito ofrecida, pues más depositantes tendrán incentivos para colocar más dinero en el banco ahora que el premio es mayor; la curva de demanda proviene de los bancos y será decreciente, pues es una demanda derivada de la demanda de crédito, dado que el banco es una empresa intermediaria.

A la tasa de interés pasiva de equilibrio, todos los depositantes que quieren transar a este precio logran hacerlo. A la tasa de interés activa de equilibrio, todos los individuos que son elegibles transan las cantidades que desean a este precio.

¿Se podría decir que el mercado de crédito es walrasiano? Cuando se considera la demanda total, el mercado de crédito no es walrasiano, pues al precio del mercado existe exceso de demanda de crédito. Pero cuando se considera la demanda efectiva solamente, el mercado de crédito es walrasiano, pues al precio del mercado el exceso de demanda es cero. El mercado de crédito, diremos, es *cuasi walrasiano.*

En un mercado walrasiano, oferta y demanda son independientes. En el mercado de crédito, en cambio, demanda y oferta no son independientes, pues los bancos determinan la demanda efectiva al decidir sobre el umbral de préstamos. Si los bancos redujeran estos umbrales,

la demanda efectiva sería mayor, aunque la demanda total de crédito sigue siendo la misma.

Según este modelo de la teoría del mercado crediticio, este funciona con exclusiones. El modelo predice que la tasa de interés activa es superior a la pasiva, que los bancos operan con colateral y que no existe un mercado de crédito al por menor. También predice que el mercado de crédito opera con exceso de demanda de crédito. Este exceso de demanda dará lugar a distintas formas de intercambio del crédito fuera del sistema bancario. Estas son las predicciones de la teoría del mercado de crédito sobre la situación de equilibrio del mercado.

Las variables exógenas del modelo son el stock de capital físico de la economía y su grado de concentración, el monto de la inversión total, el stock de dinero de la economía, la tasa de encaje y la tasa de interés internacional. Cambios en estas variables modificarán las cantidades y tasas de interés de equilibrio del mercado de crédito. Un par de proposiciones beta que son interesantes para los capítulos que vienen son las siguientes. Cuanto mayor sea el stock de capital, manteniendo el grado de concentración, mayor será la cantidad de crédito y mayor serán las tasas de interés. Para un stock de capital dado, a mayor concentración del capital —lo cual eliminaría firmas pequeñas que no llegan al umbral de elegibilidad—, mayor serán las cantidades y tasas de interés en el mercado de crédito.

¿Se parecen este mercado de crédito abstracto y el mercado de crédito del mundo real? Las regularidades del mercado de crédito del mundo real coinciden con esas predicciones: la tasa de interés activa es usualmente mayor que la tasa pasiva; los bancos operan usualmente con colaterales; no es usual observar mercados al por menor en el crédito bancario; y finalmente en el mundo real, para los pobres existen otras formas de intercambio de crédito, como cooperativas, crédito familiar, crédito informal y el crédito estatal.

El modelo de la teoría del mercado crediticio no es refutado por los datos básicos de la realidad; por lo tanto, no existe razón alguna para rechazar esta teoría en esta etapa de nuestra investigación.

Mercado de seguros

¿Por qué la gente compra seguros y por qué existen firmas que los venden? El intercambio debe ser beneficioso para ambas parte, sino no se llevaría a cabo. ¿Qué cosa lo hace beneficioso? Pero sobre todo nos interesa saber, ¿qué papel juega el mercado de seguros en la acumulación de capital físico?

Una teoría del mercado de seguros se presentará ahora. El primer supuesto es que los individuos operan en un contexto de incertidumbre. Nadie sabe con certeza las consecuencias futuras de sus acciones de hoy. Tampoco saben las sorpresas que el futuro les deparará, sean beneficiosos o perjudiciales. Se pueden distinguir cuatro contextos particulares de incertidumbre, según sean las características de los riesgos. Por el lado de la medición, los riesgos pueden ser medibles y no medibles, en el sentido de que se conocen o no se conocen las probabilidades de ocurrencia de los eventos; por el lado de la capacidad de soportabilidad que tiene el individuo, como se mostró antes, los riesgos pueden ser soportables y no soportables, en el sentido que las pérdidas puedan ser absorbidas por el individuo porque cuenta con suficiente riqueza o pueden crearle un desastre económico y social si su riqueza es insuficiente.

Tenemos así una matriz de cuatro celdas. La primera celda sería la de riesgos medibles y soportables, donde el ejemplo es el contexto que enfrentan los inversionistas, como vimos anteriormente. La segunda sería la de medibles e insoportables, donde los eventos pueden generar un desastre económico y social, y donde la gente busca transferir a otros parte del riesgo, como por ejemplo a través del mercado de seguros. La tercera sería la de no medibles y soportables, que se refiere a los avatares de la vida cotidiana y donde la gente busca sus propios mecanismos de autoprotección. Finalmente, la cuarta celda sería la de riesgos no medibles y no soportables, que se refiere a golpes fuertes, como el que proviene de los fenómenos de la naturaleza —terremotos, huracanes— y donde la gente busca protección en los bienes públicos.

El mercado de seguros existe, como se puede ver, en un contexto particular. La gente desea transferir parte del riesgo a otros y se crea una demanda de seguros. Las firmas que ofrecen seguros pueden hacer el cálculo de costos y beneficios porque las probabilidades de ocurrencia de los eventos son mensurables. En los otros contextos, el mercado de seguros es inviable.

Consideremos un modelo teórico del mercado de seguros donde se supondrá que la estructura del mercado es de competencia perfecta. La demanda proviene de los capitalistas, quienes buscan asegurar su stock de capital para eliminar o reducir el riesgo del desastre económico y social, el riesgo de salir de la clase capitalista. Al precio de mercado del bien seguro, desearán comprar una cantidad de seguros.

El precio se denomina la tasa de prima del seguro y es igual a un porcentaje por unidad de tiempo de la cantidad asegurada. Si el precio es igual a la tasa de 1% mensual, el gasto total de asegurar un stock de capital por cien mil dólares es igual a mil dólares mensuales. Al gasto total se le denomina la prima total. Si la tasa subiera a 2%, ¿cuánto quisieran asegurar los capitalistas? Necesitamos una teoría sobre el comportamiento de los capitalistas, propietarios de las firmas que producen el bien B.

Las firmas que producen el bien B buscan que maximizar las ganancias del stock de capital que poseen, pero también enfrentan riesgos de pérdida del stock de capital debido a siniestros. Considere una firma que obtiene una ganancia determinada y decide no comprar seguros. Se enfrenta ahora a una situación donde esa ganancia está acompañada de dos eventualidades: tiene una probabilidad de mantener su stock de capital y otra probabilidad de perderlo; entonces, puede tener el nivel inicial de ganancias pero con la posibilidad de perder todo su capital, y ¡dejar de ser capitalista!

Con un mercado de seguros, el capitalista enfrenta ahora las siguientes opciones: cuanto más seguro compra, y más asegurado está su stock de capital, menores serán sus ganancias netas, y cuanto mayor es su ganancia neta, más inseguro está su stock de capital. Los capitalistas

enfrentan entonces el dilema de elegir entre ganancias y seguridad de su stock de capital, o entre ganancias y permanencia en la clase capitalista.

Suponemos que la lógica del capitalista es otra. Como propusimos en el cuarto capítulo, la teoría es que los capitalistas buscan, primero, mantener su posición de clase y luego maximizar las ganancias. Para mantener su supervivencia social —no la fisiológica de Maslow—, el capitalista buscará mantener su actual stock de capital de manera indestructible. Este objetivo lo logrará comprando un seguro. Este seguro implica un costo fijo para la firma y entonces las ganancias se reducen en ese monto.

En este modelo, supondremos que los capitalistas prefieren mantener indestructible su stock de capital y entonces lo aseguran completamente, independientemente del precio del seguro. Así mantienen su status social de miembros de la clase capitalista. Con la compra del seguro, el stock de capital se vuelve indestructible, pues los riesgos de pérdida han sido eliminados y transferidos al mercado.

El gasto total de las firmas en comprar el seguro es un costo fijo. Es equivalente a un impuesto a las ganancias. Luego, la cantidad de trabajadores que maximiza la ganancia, la ganancia total y la neta del costo del seguro, seguirá siendo la misma. El costo del seguro no modificará la curva de la demanda de trabajo de las firmas que producen el bien B.

Consideremos ahora el comportamiento de la firma que produce el bien seguro. Esta firma busca maximizar las ganancias, sujeta a los costos y a las condiciones de demanda que enfrenta. Como las firmas son precio-aceptantes, la única decisión es la cantidad a producir, la cual depende de sus costos. Los costos dependen, entre otros factores, de la cantidad de siniestros que ocurran. En ese sentido, las ganancias están sujetas a riesgos, es decir, cuando ocurran pocos siniestros las ganancias serán altas y cuando ocurran muchos siniestros las ganancias serán bajas o generarán pérdidas.

El riesgo que es particular a la firma que produce seguros es el que proviene del problema de la información asimétrica. El negocio es riesgoso porque la firma no conoce bien a los clientes. Existen dos tipos de problemas de información asimétrica. El primero se llama el *problema*

de la selección adversa. Entre dos grupos de individuos, uno de alto riesgo de siniestro y el otro de bajo riesgo, ¿cuál de ellos estaría más interesado en comprar el bien seguro? Sería el grupo de mayor riesgo, pues son ellos los que más lo necesitan. En consecuencia, la firma no tendrá entre sus clientes una muestra aleatoria del universo de riesgos, sino una muestra sesgada, de más gente de alto riesgo. La ocurrencia de siniestros y el consecuente costo total para la firma serán más altos que si la firma tuviera una muestra aleatoria de clientes.

El otro problema de información asimétrica es el *problema del azar moral.* Si un individuo pierde su capital, el costo de la pérdida lo asume totalmente; si el individuo compra un seguro para su capital, el costo de la pérdida será menor, solo la parte que gasta en el seguro. El individuo tomará más precauciones en el primer caso que en el segundo. La consecuencia es que el grupo no asegurado tomará muchas precauciones para evitar los siniestros, mientras que el grupo asegurado tomará pocas precauciones y la tasa de siniestros en este grupo será mayor que en el primero. La firma tendrá un costo total mayor debido al problema del azar moral que aparece entre los asegurados.

Si el costo total de pagar por los siniestros es igual o mayor que el ingreso por las primas, la firma que produce seguros tendrá pérdidas en lugar de ganancias. Si todas las firmas de seguros enfrentan esta situación, no existirá el mercado de seguros. Si el costo total es menor que el ingreso por las primas, y es suficientemente grande como para cubrir los otros costos de la firma, habrá ganancias y existirá el mercado de seguros. En este caso, sin embargo, las firmas de seguros tienen todavía una opción adicional. Podrían tomar acciones para reducir endógenamente los riesgos de ocurrencia de los siniestros.

Estas acciones implican costos, llamados *costos de transacción.* Para reducir la selección adversa y el azar moral, la firma necesita conocer mejor a los clientes y buscar estrategias para lograr que la clientela sea mixta entre grupos de alto riesgo y bajo riesgo, así como crear incentivos positivos para que los clientes no adopten el comportamiento del azar moral, como descuentos en los precios por bajos siniestros. Todas estas

acciones implican mayores costos de transacciones. De esta manera, las firmas de seguros internalizan el problema de la información asimétrica en sus costos de producción. Si incurrir en estos costos implica mayores ganancias la firma tendrá, incentivos en tomar estas acciones.

La otra política de las firmas de seguros es reducir el costo de transacción por dólar asegurado. Al igual que en el caso de los bancos, el costo de transacción por cliente es fijo, independientemente de la cantidad de seguros transada, hasta un cierto umbral. El costo de transacción *por dólar asegurado* será entonces decreciente y luego del umbral se hará constante. Vender un seguro de tamaño pequeño (mil dólares) implica un costo igual que vender un seguro de tamaño grande (cien mil dólares); por lo tanto, el costo por dólar será relativamente alto para seguros pequeños. Luego, será mucho más rentable para la firmas de seguros vender seguros de tamaño grande en lugar de pequeños.

Las firmas de seguros establecerán un umbral del tamaño del seguro a partir del cual están dispuestas a vender. Esta política hará que los capitalistas que tengan capital cuyo valor está por debajo de este umbral quedarán excluidos del mercado de seguros. Por el lado de la demanda de seguros también podemos distinguir entre la demanda total y la demanda efectiva, constituida solo por aquellos que son elegibles por la firma de seguros para hacer la transacción.

Las curvas de demanda efectiva de seguros y de oferta de seguros determinarán el precio de equilibrio y la cantidad de equilibrio. Precios y cantidades se resolverán de manera simultánea. ¿Es este un mercado walrasiano? Si se considera la demanda total, no es un mercado walrasiano, pues al precio del mercado existe exceso de demanda. Si se considera la demanda efectiva, es un mercado walrasiano, pues al precio del mercado no existe exceso de demanda. El mercado de seguros, diremos, es cuasi walrasiano. En este mercado, la demanda no es independiente de la oferta, pues las firmas de seguros intervienen en la determinación de la demanda. Si las firmas redujeran el nivel del umbral de elegibilidad, la demanda efectiva aumentaría, a pesar de que la demanda total sigue siendo la misma.

Según este modelo, el mercado de seguros opera con exclusiones. Esta es la condición de equilibrio. El mercado de seguros no opera con transacciones al por menor. Y estas predicciones son consistentes con las regularidades que observamos en el mundo real. Así, no se puede entrar a un supermercado y comprar la cantidad de seguro que uno desee al precio del mercado. La gente pobre utiliza formas distintas a las del mercado de seguros, desde las redes familiares hasta programas estatales, para protegerse de los riesgos.

La variable exógena del modelo es el stock de capital físico. Una proposición beta del modelo dice que el tamaño del mercado de seguros depende no solo del tamaño del stock de capital físico de la economía, sino también de su grado de concentración. Cuanto menos concentrado sea, el mercado de seguros será menos desarrollado, pues la mayoría de los capitalistas no pasarán el umbral de stock de capital necesario para acceder al mercado de seguros. Ciertamente, esta conclusión es similar a la del mercado de crédito.

Algunas conclusiones comunes para ambos mercados —crédito y seguro— merecen destacarse. Primero, los datos de la realidad del tercer mundo son consistentes con las predicciones comunes: son mercados que excluyen a los pobres. Los millones de campesinos que existen en cada país, y que en el agregado suman una enorme cantidad de tierras, no tienen acceso ni al mercado de crédito ni al mercado de seguros. Los millones de pequeñas unidades urbanas que existen en cada país, y que en el agregado suman cantidades importantes de capital físico, tampoco tienen acceso ni al mercado de crédito ni al mercado de seguros.

Segundo, los dos mercados juegan un papel importante en la acumulación del capital físico. Para acceder a ambos mercados, y así poder obtener capital físico y mantenerlo indestructible, el individuo debe, primero, estar dotado de capital físico y en una cantidad significativa. El funcionamiento de ambos mercados, con exclusiones de los individuos sin capital o poco capital, constituye una limitación para que los trabajadores puedan llegar a ser capitalistas. Esta predicción de la teoría es consistente con la realidad: la reproducción de las clases sociales.

Ambos modelos suponen competencia perfecta. Se puede decir que en el mundo real se observa que los mercados de crédito y seguros están dominados por pocas firmas, todas de gran tamaño. Parecería así «poco realista» suponer una competencia perfecta. Si se cambiara el supuesto de competencia perfecta por otra estructura de mercado, como monopolio, cartel, competencia imperfecta, el modelo sería un poco más complejo. Pero, y aquí lo más importante, se puede mostrar que con esos modelos se llegará a las mismas predicciones empíricas que hemos obtenido con el modelo de competencia perfecta.

Las razones son fáciles de señalar. La lógica de las firmas seguirían siendo las mismas. Los precios y cantidades seguirían siendo endógenos. Contra todo lo que se cree, cuando existe poder de mercado, la firma no fijará «el precio que le da la gana», como si fuera un precio arbitrario y exógeno al proceso económico; la firma con poder fijará el precio que más le conviene, es decir, el que hace máxima la ganancia. Debido a que debe interactuar en el mercado con los compradores, este precio seguirá siendo endógeno.

Ciertamente, distintos modelos generarán distintas predicciones empíricas sobre otros aspectos del comportamiento de los mercados, pero sobre los comportamientos que nos han interesado establecer aquí, como el equilibrio con exclusión, todos los modelos llegarán a las mismas predicciones. Y estas predicciones son, como hemos visto, consistentes con las regularidades que observamos en el mundo real. Por lo tanto, no existe razón alguna para rechazar los modelos de las teorías de los mercados de crédito y de seguros que hemos presentado y entonces podemos aceptar ambas teorías en esta etapa de la investigación.

Conclusiones

En este capítulo hemos explicado el comportamiento de los inversionistas. La teoría que hemos desarrollado para explicar la asignación de las inversiones mundiales entre el primer mundo y el tercer mundo

predice, que esta asignación depende de las diferencias en la dotación factorial y en la desigualdad inicial entre ellas. Así hemos podido explicar el hecho empírico de que la mayor proporción de la inversión mundial se va al primer mundo.

Luego, hemos desarrollado en este capítulo teorías sobre el funcionamiento del mercado de crédito y de seguros. Estas teorías predicen que estos mercados operan con exclusiones. Las regularidades en el funcionamiento de estos mercados en el mundo real son consistentes con estas predicciones. Estas teorías nos permiten explicar por qué los trabajadores no pueden llegar a acumular capital físico y, así, llegar a ser capitalistas. El funcionamiento de los mercados de crédito y seguros constituyen una limitación para la movilidad social. Estas teorías, en suma, nos permiten explicar una de las regularidades fundamentales del capitalismo: la reproducción de la desigualdad y de la estructura de clases.

Hasta aquí hemos mostrado que existen tres mercados que funcionan con exclusiones. Este es el caso de los mercados de crédito y seguros, que hemos estudiado en este capítulo, y también es el caso del mercado laboral, que fue estudiado en el cuarto capítulo. A estos tres mercados que operan con exclusiones podemos denominarlos *mercados básicos*, pues juegan un papel central en la reproducción de la desigualdad y de las clases sociales.

Hemos seleccionado estos mercados como los mercados esenciales para explicar la producción y distribución en el sistema capitalista. Hemos supuesto que estos mercados son no-walrasianos y cuasi-walrasianos, es decir, son mercados que operan con exclusiones, y en particular excluyendo a los pobres. Los otros mercados también tienen efectos, pero hemos supuesto que no tienen tanta importancia y por eso los hemos ignorado. No todos los mercados tienen la misma importancia en el sistema capitalista, unos son más importantes que otros, y los mercados de trabajo, de crédito y de seguros son los esenciales o los mercados básicos. Esta es la teoría de los mercados que hemos desarrollado hasta aquí. Las predicciones de esta teoría de los mercados son consistentes, hasta ahora, con los datos de la realidad.

Si la inversión en capital físico no abre posibilidades a la movilidad social en el sistema capitalista, ¿qué podemos decir sobre la inversión en capital humano?, ¿es la educación un sistema igualador? La respuesta será desarrollada en el siguiente capítulo.

Capítulo 7

Acumulación del capital humano

Se denomina capital humano al stock de conocimientos productivos que están incorporados en el individuo. Capital humano, como todo capital, es un stock y como todo stock puede ser acumulado. Según la literatura económica, el capital humano juega un papel importante en el proceso de crecimiento económico de las sociedades.

El capital humano ha sido considerado hasta ahora una variable exógena en los modelos teóricos presentados, pero ahora será transformado en una variable endógena. ¿Cuáles son los factores que determinan el capital humano de los individuos y de la sociedad?, ¿qué efectos tiene la acumulación del capital humano en la producción y distribución? Este capítulo tiene por objeto responder estas preguntas. Para ello se presenta una teoría del capital humano que intentará explicar las causas y consecuencias de su acumulación.

Transformación de educación en capital humano

Ciertamente, la gente no nace con su capital humano. Necesita invertir para adquirirlo, y lo hace a través del proceso educativo. Tampoco el capital humano se construye sobre tabula rasa. En la literatura científica que trata de explicar los procesos de aprendizaje —como psicología, biología, neurociencia—, las dotaciones intelectuales iniciales que trae consigo el individuo al proceso educativo son fundamentales para el

aprendizaje. Sobre estas dotaciones iniciales, la teoría estándar en esa literatura sostiene que los talentos humanos son múltiples, la llamada teoría de la inteligencia múltiple (Gardner 1999). Pero, si bien nuestros talentos son diversificados, su composición no es homogénea entre individuos, pues unos tienen mayor talento para el arte, otros para la ciencia, etcétera. Por consiguiente, los humanos somos diferentes en nuestras dotaciones intelectuales iniciales.

En esa literatura también se supone que, si bien las dotaciones iniciales de talentos del individuo, su herencia genética, son exógenas —el efecto de *nature*—, los talentos no están dados de una vez para siempre. Los talentos también pueden ser desarrollados a lo largo del tiempo. Dado los talentos iniciales, los talentos actuales del individuo son entonces endógenos y dependen de su medio social —el efecto de *nurture*—. Como ha mostrado la teoría de la plasticidad del cerebro en la literatura moderna de la neurociencia,

> El cerebro no es una computadora que simplemente ejecuta programas genéticamente predeterminados. Tampoco es una calabaza pasiva, víctima de las influencias del medio que recaen sobre ella. Genes y ambiente interactúan para modificar continuamente nuestro cerebro, desde el momento en que somos concebidos hasta el momento de nuestra muerte (Ratey 2002: 17).

Los talentos individuales que provienen de la genética se pueden suponer que son aleatorios, pero no así los que provienen del medio social. La célebre distinción que hizo Rousseau (1755) se refiere justamente a estos factores. Existen, dijo Rosseau, dos tipos de desigualdades entre los individuos: las *naturales,* que son las dotaciones naturales, regalos de la naturaleza y que son aleatorias; y las *artificiales,* que se originan en el funcionamiento de la sociedad.

En el proceso educativo, los talentos de los individuos, vistos ahora como capacidades de aprendizaje, se combinarán con los insumos que provienen de la escuela para producir conocimiento general y capital humano. Aquí se considerará que conocimiento general y capital

humano no son independientes; por el contrario, se supondrá que un mayor nivel de capital humano requiere un mayor nivel de conocimiento general. También debe quedar claro que el concepto de «escuela» utilizado aquí se refiere a todos los niveles educativos, tales como primaria, secundaria, técnica y universitaria.

Se presentará ahora una teoría del proceso educativo. El supuesto primario es que en las sociedades capitalistas el proceso educativo no es uniforme, sino diferenciado y jerarquizado por clases sociales. De los dos factores que determinan los talentos de los individuos, el factor esencial es la clase social a la cual pertenecen los individuos. El factor natural o genético, siendo un factor aleatorio, se le considerará de menor importancia, especialmente cuando el análisis se hace en términos de clases o grupos sociales. La escuela también es diferenciada y jerarquizada por clases sociales.

Un modelo de la teoría será construido ahora incluyendo supuestos auxiliares. Supondremos que existen tres sociedades capitalistas —épsilon, omega y sigma— y los grupos sociales dentro de cada una, tal como fueron definidos anteriormente; es decir, épsilon y omega se componen de los grupos sociales A-Y y sigma se compone de A-Y-Z. Las tres son sociedades desiguales y entonces los individuos participarán en el proceso educativo dotados de sus capacidades de aprendizaje que están desigualmente distribuidas. Los mecanismos por los cuales los ricos pueden generar mayores capacidades de aprendizaje en sus hijos, comparado al de los pobres, incluyen nutrición, salud, estimulación intelectual temprana y lenguaje.

La nutrición tiene un efecto directo sobre el cerebro y la capacidad cognoscitiva del individuo (Ratey 2002); tiene, además, un efecto indirecto vía los episodios de enfermedades. Y el grado de nutrición depende positivamente del ingreso de las familias.

El estado de salud de los individuos tampoco es neutro con respecto a la desigualdad. Así, en promedio, los episodios de enfermedades serán menos frecuentes en los hijos de familias ricas; además, para su tratamiento, las familias ricas utilizan servicios de salud en distinta calidad

y cantidad, dependiendo de las diferencias en el ingreso real y en la provisión del servicio de salud del mercado y del sector público.

En una sociedad épsilon, el factor más importante es el ingreso real, pues el acceso a los servicios públicos de salud opera como bienes públicos universales, dado el grado uniforme de ciudadanía. En una sociedad sigma, por contraste, los dos factores cuentan, pues aquí existen tres tipos de servicios de salud: los privados, los públicos de buena calidad y los públicos de mala calidad. Los pobres acceden a estos últimos. Dada la diferencia en grados de ciudadanía, los servicios públicos de salud se ofrecerán como bienes públicos locales —diferenciados en calidad y derecho de acceso por grupos sociales—.

Las diferencias en el estado de salud también vienen del lado de la salud ambiental, que tampoco es neutral con respecto a la desigualdad. Las familias ricas pueden evitar los problemas de saneamiento ambiental —calidad de agua, calidad de desagüe y contaminación del aire— mediante la «salida», pues para eso pueden construir barrios residenciales exclusivos. Las familias pobres solo pueden intentar resolver el problema con su «voz»: solicitar, protestar, etcétera.

Se supondrá que la estimulación intelectual temprana de los hijos también depende del nivel socioeconómico de los padres. La mayor cantidad, calidad y diversidad de bienes y servicios que consumen las familias ricas inducen a descubrir los talentos de sus hijos y a su mayor desarrollo intelectual.

Otro factor de desigualdad en la capacidad de aprendizaje que se origina en el nivel socioeconómico de los padres será el lenguaje. Existen desigualdades lingüísticas entre individuos. Estas desigualdades se manifiestan en varios aspectos del uso del lenguaje, tales como vocabulario, sintaxis, formas de hablar, capacidades de lectura y escritura. Según la teoría sociolingüística, las desigualdades en el lenguaje se originan más en las experiencias del individuo —su medio social— antes que en factores genéticos (cf. Hudson 1996: 204).

La importancia del problema de las desigualdades en el lenguaje será mayor en sigma que en épsilon. Sociedades multiculturales son

también multilingües; y dada las jerarquías sociales que existen, unas lenguas dominan a otras. La lengua del grupo social dominante será también la lengua dominante en la sociedad. Así la disparidad lingüística se convierte en desigualdad lingüística. Por otro lado, dada la segregación que existe en sigma, la lengua dominante no es uniforme. Así aparece el fenómeno de la *heteroglosia*, que es otra forma de desigualdad lingüística.

Heteroglosia es un concepto que se utiliza en la teoría sociolingüística para referirse a la existencia de varias formas o variaciones de una misma lengua, con una jerarquía entre esas formas. Sigma es entonces una sociedad heteroglósica. En el uso del lenguaje dominante se puede reconocer las formas correctas y socialmente superiores —el uso que hace la población dominante de su lengua nativa— hasta las formas incorrectas y socialmente bajas —el uso que hace la población subordinada de la lengua dominante—.[5]

La desigualdad en el lenguaje entre individuos lleva a la desigualdad en la capacidad para el aprendizaje. Dos razones importantes son el problema de la heteroglosia y el de la cultura oral. La primera ya fue presentada arriba y tiene que ver con las limitaciones que tienen las poblaciones subordinadas en el manejo de la lengua dominante. Sobre la segunda, la teoría es que los pensamientos abstractos y complejos requieren símbolos, como en el caso de las matemáticas; y en términos lingüísticos, requieren del leguaje escrito (Searle 1995). La implicancia de esta teoría es que las sociedades que utilizan solo un leguaje oral están en desventaja en cuanto a su capacidad de aprendizaje frente a las sociedades con cultura de lenguaje escrito.

---

[5] Las diversas formas del castellano que se habla en el Perú, por ejemplo, incluye «castellano limeño», «castellano estándar», «castellano costeño», «castellano serrano», «castellano charapa», «castellano motoso», donde los dos primeros tienen mayor jerarquía social. La falta de comando en el castellano de las poblaciones de legado aborigen se nota, aparte del acento al hablarlo, en el uso limitado de oraciones que expresen razonamiento abstracto, lo cual requiere el uso de verbos en subjuntivos e impersonales. La cultura oral de estas poblaciones debe estar influyendo en este problema lingüístico.

En un libro que ya es un clásico, Jared Diamond (1999), un biólogo americano, se pregunta, ¿por qué los españoles conquistaron a los incas y aztecas y no fue al revés? Su respuesta es que un factor importante fue el lenguaje escrito. Entre los factores iniciales que explican la superioridad de occidente en el desarrollo humano se encuentra, según Diamond, la cultura del lenguaje escrito.

Considere por un momento una sociedad épsilon en donde la gente tiene la cultura del leguaje escrito y nadie es analfabeto. Por contraste, considere también una sociedad sigma donde los trabajadores Y viven en un ambiente donde prima la cultura del lenguaje escrito y nadie es analfabeto, y donde los trabajadores Z viven en ambiente donde prima la cultura oral (su lengua nativa no es escrita) y son analfabetos en la lengua dominante. En este contexto, existirá desigualdad lingüística y también desigualdad en la capacidad cognitiva entre las sociedad épsilon y sigma, así como dentro de sigma.

En la sociedad sigma, los hijos de la población Z tendrán limitaciones para el aprendizaje de pensamientos abstractos y complejos debido en parte al analfabetismo del medio y en parte a la cultura oral. Ambos efectos se refuerzan y reducen su capacidad cognitiva. Por lo tanto, los hijos de la población Z tendrán, en promedio, una capacidad cognitiva inferior en comparación a los de la población Y. Esta desigualdad existirá aun entre familias Y y familias Z que son analfabetas, pues provenir de una familia analfabeta dentro de una cultura escrita es menos limitante para el aprendizaje que provenir de una de cultura oral. Esta desigualdad existirá incluso entre familias Y y familias Z que no son analfabetas en la lengua dominante, debido al efecto de la cultura oral. La desigualdad en la capacidad cognitiva de los individuos será menos significativa en las sociedades épsilon y omega, que son socialmente homogéneas.[6]

[6] Aun al final del proceso educativo, la desigualdad lingüística se mantendrá en una sociedad sigma. La lengua es, en esta sociedad, un marcador social: «Déjame escuchar cómo hablas y te diré quién eres». Este marcador se refiere naturalmente a la lengua dominante.

Desigualdades lingüísticas implican desigualdades en capacidades de aprendizaje o cognitivas. Una hipótesis empírica que viene de la teoría sociolingüística sobre las relaciones entre desigualdad social y lingüística va mucho más allá:

> La desigualdad lingüística puede ser vista como una *causa* de la desigualdad social, pero también como una *consecuencia*, debido a que el lenguaje es uno de los medios más importantes a través de los cuales la desigualdad social se reproduce de generación en generación (Hudson 1996: 205).

En suma, la teoría del proceso educativo supone que, en cada sociedad, los individuos que participan en el proceso educativo están dotados de capacidades cognitivas desiguales, que se asocian a la desigualdad de ingresos, la cual a su vez se origina en la desigualdad inicial en la distribución de activos económicos y sociales. El modelo de la teoría que presentamos aquí utiliza el supuesto auxiliar de que el proceso educativo toma lugar en las sociedades épsilon, omega y sigma que conocemos. Esta proposición es teórica, pues las capacidades cognitivas de los individuos no son observables. Bajo estas condiciones, la gente busca acumular capital humano. En el proceso educativo se combinarán las capacidades cognitivas de los individuos con los insumos de la escuela para producir conocimiento general y capital humano.

¿Cómo se puede explicar la desigualdad en capital humano que existe entre grupos sociales? Otro supuesto del modelo es que el proceso educativo escolarizado es el mecanismo esencial para adquirir capital humano. Otras formas, como cursos no escolarizados y capacitación dentro de las firmas, serán consideradas menos importantes y serán, por eso, ignoradas.

Suponga por el momento que las escuelas son todas homogéneas. Aun bajo estas condiciones, y para una misma cantidad de años de escolaridad, el proceso educativo generará diferencias en el capital humano entre los hijos de los ricos y de los pobres, debido a las desigualdades en las capacidades cognitivas.

Pero no todos tendrán los mismos años de escolaridad. También aquí habrá un efecto de la desigualdad económica. Al igual que en el caso del capital físico, la acumulación de capital humano requiere financiamiento. Los ricos tendrán una mayor capacidad de financiamiento que los pobres. El efecto ingreso es positivo: la cantidad demandada de capital humano dependerá positivamente del nivel de ingreso de las familias, el cual llevará a que individuos que provienen de familias ricas tengan mayor número de años de escolaridad que los que provienen de familias pobres.

Pero no todas las escuelas son iguales. Si se abandona el supuesto de que las escuelas son homogéneas, existirá otro factor de desigualdad. Si la escuela privada es más equipada y cuenta con mejores profesores que la escuela pública, o la escuela urbana es de mejor calidad que la escuela rural, la acumulación de capital humano depende del tipo de escuela al que asiste el individuo. Esta diferencia llevará naturalmente a que los hijos de los ricos asistan a escuelas de mejor calidad.

Así como existe una tecnología para producir máquinas, también se puede suponer la existencia de una tecnología para producir capital humano. Luego, se puede ver la acumulación de capital humano como un proceso productivo, cuyo producto es la cantidad de capital humano que produce una escuela y que dependerá de los insumos que ingresen al proceso. Según los supuestos presentados anteriormente, los insumos se pueden clasificar en tres: la calidad de los estudiantes —capacidad de aprendizaje—, la calidad de la escuela —cantidad y calidad de equipamiento y cantidad y calidad de profesores— y la duración del proceso —años de escolaridad—. Estos insumos no son uniformes para los individuos, sino que son distintos para distintos grupos sociales.

Los hijos de familias ricas tendrán, en promedio, más capital humano que los de familias pobres debido a tres efectos:

(a) capacidad de aprendizaje: a igualdad de años de escolaridad, y de calidad de escuela, los hijos de los ricos tendrán un mayor nivel de capital humano;

(b) ingreso: tendrán más años de escolaridad;
(c) de la escuela: asistirán a escuelas de mejor calidad.

Los tres efectos se originan en la desigualdad de ingresos entre las familias. Pero esta desigualdad en los flujos de ingresos se origina en la desigualdad inicial en la dotación de activos económicos y sociales. Luego, los tres efectos provienen de las desigualdades en el medio social en el que viven los individuos.

Esta teoría del proceso educativo supone, en suma, que educación no es igual que capital humano; y que la educación es el proceso que transforma insumos en el producto capital humano. Aunque el proceso productivo se lleva a cabo en la escuela, los insumos no incluyen solo la calidad de la escuela, sino también otros que vienen de fuera de la escuela, como la calidad del estudiante.

Esta teoría es general pues se aplica a los tres tipos de sociedades capitalistas. Para obtener proposiciones beta se pueden utilizar los modelos particulares de estas sociedades desarrolladas en el cuarto capítulo.

Considere el modelo simple de la sociedad sigma, donde la estructura social se compone de tres grupos sociales jerarquizados A-Y-Z. En sigma existirá una relación positiva entre años de escolaridad y nivel de capital humano. Pero esta relación no será unívoca para los tres grupos sociales sino que tomará formas particulares para cada grupo social. La razón está en que los hijos del grupo Z asistirán a un proceso educativo donde la calidad del estudiante y la calidad de la escuela son menores que en el caso de los otros grupos sociales, debido a que el grupo Z está muy poco dotado de activos económicos y políticos. Su acceso a la escuela pública será a la de baja calidad. En el siguiente orden se encuentra el grupo Y. Los gobiernos ofrecen una educación segmentada, según los grupos sociales. Es así como funciona la democracia en una sociedad sigma.

El proceso educativo en la sociedad sigma se representa en la figura 7.1, en el panel (a). En el eje horizontal se mide años de educación y

en el eje vertical el nivel de capital humano. Las tres líneas escalonadas representan la transformación de educación en capital humano para cada grupo social A, Y, Z. Para un mismo número de años de escolaridad, el nivel de capital humano será mayor en el grupo A, le seguirá el del grupo Y y el más bajo le corresponderá al grupo Z. A mayores años de escolaridad, el nivel de capital humano aumentará en los tres casos, pero cada grupo se moverá a lo largo de una curva diferente.

La consecuencia de esta jerarquía en el proceso educativo es que si el grupo Z quisiera tener el mismo nivel de capital humano de A, tendría que tener más años de escolaridad. Eso es inviable, pues debido a su restricción económica el grupo social Z tendrá, por el contrario, una cantidad menor de años de escolaridad que el grupo A o el grupo Y. No podrán igualarse los niveles de capital humano entre los hijos de los tres grupos sociales que conforman la sociedad sigma. El proceso educativo no es igualador. Este es el resultado de que el proceso educativo opera en una sociedad con una desigualdad inicial muy pronunciada.

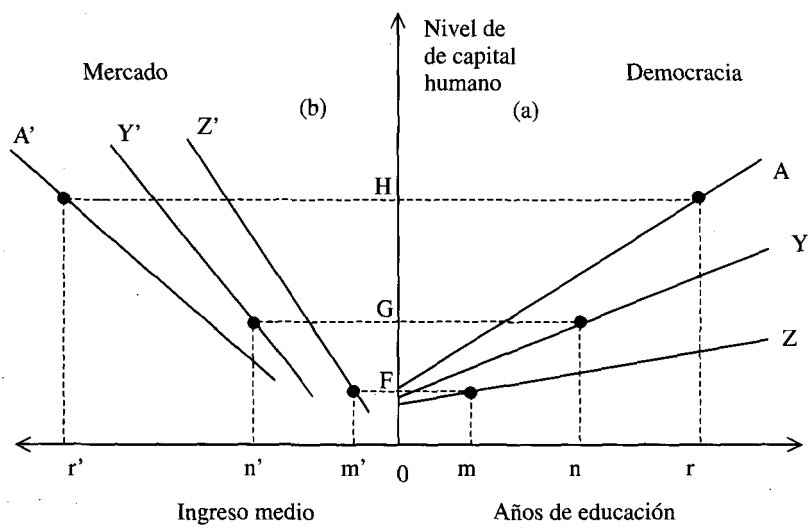

Figura 7.1 Relaciones entre educación, capital humano e ingresos medios

En el caso de las sociedad épsilon, donde el grupo social Z no existe, las relaciones entre años de escolaridad y nivel de capital humano mostrarán solo dos curvas, también jerarquizadas: la superior que representa el caso del grupo social A y el inferior del grupo social Y. En esta sociedad tampoco el proceso educativo es igualador, pero las brechas no serán tan grandes como en sigma, pues las desigualdades iniciales tampoco son tan pronunciadas como en sigma. Habrá, ciertamente, diferencias cuantitativas en los resultados, aunque cualitativamente la conclusión es la misma: la desigualdad inicial juega un papel central en la desigualdad en la acumulación de capital humano. El caso de la sociedad omega será similar al de épsilon. Los procesos educativos de las sociedades épsilon y omega también se pueden representar en el panel (a) de la figura 7.1 si se ignora la línea Z.

Las predicciones empíricas de este modelo de la teoría del proceso educativo son tres. Primero, a más años de escolaridad le corresponde mayor nivel de capital humano en todos los grupos sociales. Segundo, la relación años de escolaridad y nivel de capital humano no es la misma para todos los grupos sociales; la curva que los relaciona tiene un nivel más elevado para los ricos que para los pobres. Tercero, los ricos tienen más años de escolaridad. En suma, la desigualdad en la acumulación de capital humano sigue a la desigualdad inicial; es decir, el proceso educativo no es igualador de capital humano.

Transformación del capital humano en ingresos

La transformación de capital humano en ingresos operará a través del mercado laboral. Para explicar esta relación se utilizarán los modelos de la teoría del mercado laboral que fueron desarrolladas en el cuarto capítulo.

A mayor capital humano de un grupo social le corresponderá un mayor nivel de ingresos. Esta relación simplemente refleja la rentabilidad positiva de la inversión en capital humano. Esta rentabilidad obedece

al efecto positivo que tiene el capital humano sobre la productividad laboral. Este efecto opera por dos vías. La primera se debe a la complementariedad entre el capital humano y el capital físico, pues para unas máquinas dadas, la cantidad de producto obtenido dependerá del capital humano de los trabajadores. Si su capital humano aumenta, los trabajadores le darán un uso más productivo a esas máquinas y obtendrán una mayor cantidad de producto. Por ejemplo, saber leer correctamente las instrucciones para operar una máquina significará darle usos más productivos en comparación a no saber leer y no conocer bien el manejo numérico.

La segunda se debe a la complementariedad entre nuevas tecnologías y capital humano. Las nuevas tecnologías están incorporadas en las máquinas modernas, cuya operación requiere mayor capital humano. El trabajador no podría operarla con eficiencia si mantuviera su mismo capital humano. Se necesita un umbral de capital humano para operar con tecnologías modernas. Por ejemplo, al introducirse las computadoras en la producción, los trabajadores tuvieron que ir a un proceso de aprendizaje para saber operarlas y surgieron las escuelas técnicas de pos secundaria y las escuelas de ingeniería electrónica en las universidades.

Ciertamente, el capital humano puede también acumularse en la misma empresa. A este caso se le llama entrenamiento o capacitación en el mismo lugar de trabajo —llamado usualmente *on the job training*—. En este caso, la firma financia la acumulación de capital. En este estudio ignoraremos este caso y supondremos que el factor esencial en la acumulación de capital se da en la escuela.

El aumento en el capital humano tendrá el efecto de elevar la productividad promedio y la productividad marginal del trabajo. Según los modelos de la teoría del mercado laboral del cuarto capítulo, la demanda de trabajo aumentará y el salario real aumentará. Sería equivalente a un aumento en el stock de capital físico, cuyo efecto se estudió en ese capítulo. Luego, a mayor capital humano le corresponderá un mayor salario real. Pero en las tres sociedades el mercado laboral opera con exclusiones y entonces queda la pregunta sobre el efecto del

aumento en el capital humano sobre los ingresos en los sectores de subsistencia.

En el caso de la sociedad sigma, donde existen tres grupos sociales A-Y-Z, podemos analizar los efectos en dos etapas. Primero, si los tres grupos tuvieran el mismo nivel de capital humano, el orden de los ingresos serían los señalados en la jerarquía A-Y-Z, pues es esta situación la que supusimos en el modelo sigma, con la excepción de que los trabajadores Z tenían un capital humano inferior. Si ahora tuvieran un capital humano similar al de los trabajadores Y, se puede todavía suponer que el ingreso medio sigue siendo el más bajo de la sociedad, pues su acceso a los otros factores de producción pueden seguir siendo limitado —como calidad de tierra, distancia a mercados y a infraestructura social—.

Existen dos razones para que el mismo capital humano no genere el mismo ingreso en los trabajadores. La primera se refiere al funcionamiento del mercado laboral. Este mercado opera con exclusiones y genera sectores de autoempleo, llamados los sectores de subsistencia. Existen diferencias en la productividad laboral entre el sector capitalista y los sectores de subsistencia, que se originan en las bajas dotaciones de capital físico en los sectores de subsistencia (mostrada en el capítulo 4). La otra razón se refiere, al funcionamiento de los mercados de crédito y seguros. Estos mercados operan también con exclusiones. La consecuencia es que los trabajadores de los sectores de subsistencia no pueden elevar su productividad acudiendo a estos mercados para acumular capital físico (mostrada en el sexto capítulo). Los mercados básicos juegan así un papel importante en el proceso de transformar capital humano en ingresos.

Si los tres grupos sociales tuvieran el mismo nivel de capital humano, el grupo A tendría el ingreso medio más alto, pues este incluye el salario implícito y la ganancia capitalista; el grupo Y se ubicaría a continuación, pues su ingreso medio sería un promedio del salario y del ingreso de autoempleo, que es inferior al salario; y el grupo Z estaría en último lugar, pues su ingreso medio sería un promedio entre el

salario y el ingreso de autoempleo, que sería inferior al que se genera en el sector de subsistencia Y.

Como este resultado es válido para cualquier nivel de capital humano, tendríamos una relación particular entre capital humano e ingresos para cada grupo social. Dado que a mayor nivel de capital humano le corresponderá un mayor nivel de ingresos para cada grupo social, tendremos en conjunto tres curvas escalonadas en el orden A-Y-Z. Estas relaciones se representan en la figura 7.1, en el panel b. Allí se puede ver cómo el mercado transforma el capital humano, que se mide en el eje vertical, en ingresos, que se mide en el eje horizontal, para cada grupo social por separado a lo largo de las líneas A', Y', Z'.

Segundo, sabemos por el resultado de la sección anterior que los tres grupos no pueden tener el mismo nivel de capital humano. Como el grupo social A tiene mayor nivel de capital humano, el grupo Z tiene el menor nivel y el grupo social Y está en el medio, no existe una tendencia clara a la reducción de la desigualdad de ingresos como resultado de la acumulación del capital humano.

Las conclusiones obtenidas en la sociedad sigma se aplican también al caso de la sociedad épsilon y omega. Aunque allí existen solo dos grupos sociales, A-Y, el proceso de transformación de capital humano en ingresos es el mismo. El panel b de la figura 7.1 puede representar estas sociedades si se ignora la línea Z'.

En suma, las diferencias de ingresos entre los grupos sociales es el resultado de dos efectos. El primero supone que todos tienen el mismo nivel de capital humano. La diferencia de ingresos se deberá a las diferencias en la dotación inicial de activos económicos. Segundo, los que tienen más dotaciones de activos logran acumular mayor capital humano, lo cual da lugar a una diferencia adicional en los ingresos. Los ingresos siguen, así, la desigualdad inicial en la distribución de activos. Por lo tanto, las predicciones que se derivan de los modelos de la teoría del mercado laboral, y de los de la teoría de los mercados básicos en general, son que, en el proceso de transformación del capital humano en ingresos, el mecanismo del mercado no es igualador de ingresos.

Transformación de educación en ingresos

La tercera relación es muy sencilla, pues se deriva directamente de las dos anteriores. Es la forma resumida o reducida de las relaciones anteriores. Si en una sociedad dada, para cada grupo social existe una relación positiva entre años de educación y nivel de capital humano, y luego otra relación positiva entre nivel de capital humano y nivel de ingresos, entonces existirá una relación también positiva entre años de educación y nivel de ingresos.

También queda claro que para el caso de la sociedad sigma, esta relación aparecerá representada en tres curvas escalonadas que corresponden a los grupos sociales A-Y-Z. La forma reducida de las relaciones mostradas en la figura 7.1 aparece ahora en la figura 7.2. En el eje horizontal se mide años de educación (E) y en el eje vertical ingresos reales (y). Amigo lector: diviértase ahora obteniendo este nuevo gráfico a partir de los dos gráficos anteriores. En la sociedad épsilon se dará cualitativamente el mismo proceso, pero existirán solo dos curvas escalonadas, una para el grupo social A y otra debajo de ella para el grupo social Y. En la sociedad omega, el resultado será más parecido al de épsilon.

Tenemos así un modelo de la teoría del capital humano, construido sobre la base de dos procesos, uno sobre el proceso educativo y el otro sobre los mercados. Las relaciones establecidas hasta aquí nos indican las relaciones estructurales de la sociedad en el proceso de la acumulación de capital humano. ¿Cuál será la situación de equilibrio? Ahora tenemos que retornar a nuestros viejos amigos, los modelos de equilibrio general, e introducir este modelo de la teoría del capital humano en un modelo de equilibrio general. Esto lo haremos para cada sociedad capitalista.

Como estamos trabajando con un modelo estático de la economía, las variables endógenas son el ingreso nacional y su distribución entre los grupos sociales; y las variables exógenas son las condiciones iniciales de la economía, que incluye la dotación de factores productivos y la desigualdad inicial. Sabemos que, dada las variables exógenas, existe un

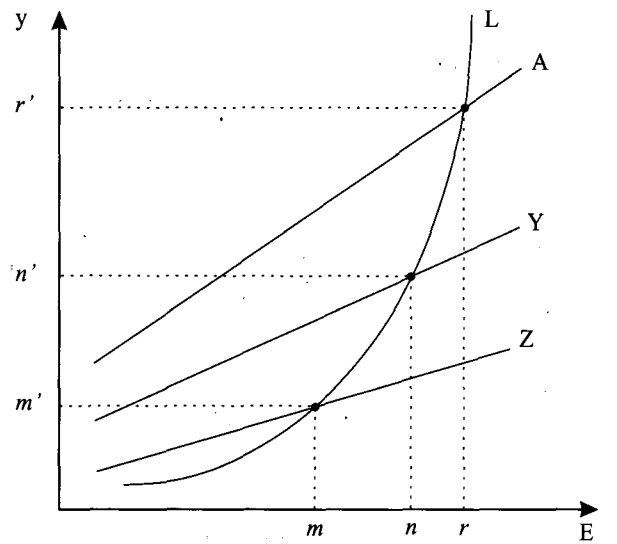

Figura 7.2 Hipótesis sobre las relaciones entre educación, ingresos y grupos sociales

equilibrio general en la economía que determina el ingreso nacional y su distribución. Así lo hemos mostrado en el cuarto capítulo para los tres tipos de sociedades capitalistas.

Pero ahora introducimos el stock de capital humano y su distribución entre los grupos sociales como nuevas variables exógenas. Dada la relación que existe entre capital humano y años de educación, mediremos el stock de capital por la variable años de educación. Luego, entre las variables exógenas se encuentra los años de educación y su distribución entre los grupos sociales.

Consideremos la sociedad sigma. Como condición inicial supondremos que los años de educación de la población A son mayores a los de la población Y, y que estas son mayores a las de la población Z. Luego, en el equilibrio general, los ingresos medios de cada grupo social seguirán ese orden. En la figura 7.2, las condiciones iniciales en la sociedad sigma se representan por los puntos m, n, r en el eje horizontal, que mide años de educación, y por m', n', r' en el eje vertical, que mide

ingresos. La línea L conecta estos puntos y es observable. Debido a que cada par de valores de años de educación e ingresos se encuentran en diferentes curvas, la desigualdad en los ingresos será pronunciada, será más que proporcional a las diferencias en los años de educación. Si el grupo social A tiene, en promedio, quince años de educación y el grupo social Z tiene, en promedio, cinco años de educación, la diferencia en el ingreso promedio entre estos dos grupos será en más del triple.

En omega y épsilon, la situación de equilibrio inicial será como en sigma, pero sin la existencia de la población Z. El equilibrio general también tendrá la misma característica: que la diferencia de ingresos es más que proporcional a las diferencias en años de educación.

La pregunta que el modelo estático puede responder se refiere al efecto de aumentos exógenos en el stock del capital humano —medido por los años de educación— sobre el ingreso nacional y su distribución (tal como analizamos el efecto de aumentos exógenos en el stock del capital físico en los modelos estáticos del cuarto capítulo). En particular, una pregunta interesante es la siguiente, ¿cuál será el efecto de una reducción en las diferencias en años de educación entre los grupos sociales sobre sus diferencias en los ingresos?

La respuesta es relativamente sencilla. Considere el caso de la sociedad sigma. Más años de educación para la población Z implicará aumentar sus ingresos a lo largo de una curva particular, que se encuentra en un nivel más bajo con relación a la de los otros grupos. La desigualdad se reduciría. Pero si los años de educación también aumentan para el grupo Y —aunque en una menor cantidad—, el cambio en la desigualdad se vuelve ambiguo. La tasa de retorno de la educación es mayor para el grupo Y que para el grupo Z, pero además el ingreso del grupo A aumenta, ya que las ganancias aumentarán con el incremento en la productividad laboral, que trae consigo la mayor educación del grupo Y.

Otro sería el resultado si la sociedad sigma tuviera una única curva para los tres grupos, como la línea A, a lo largo de la cual la población Z avanzara en ingresos con sus mayores años educativos; en este caso, la

reducción de años de educación llevaría a la reducción en las diferencias de ingresos y hasta la igualación de ingresos medios entre los grupos sociales. La educación sería igualadora de ingresos. Pero este resultado no es viable, pues la sociedad sigma no funciona así. La misma conclusión se obtendrá en el caso de la sociedad épsilon y también en omega.

Una predicción del modelo de equilibrio general estático de la teoría sigma, cuando se le introduce la teoría del capital humano, es que en el proceso de la acumulación de capital humano, y dada la desigualdad inicial, el mercado y la democracia, las dos instituciones básicas del capitalismo, no conducen a la sociedad a la igualdad en los ingresos, ni tampoco a la reducción en la desigualdad. Si bien cuando el número de años de educación aumenta, el ingreso aumentará, es decir, existe una curva creciente que relaciona años de educación con ingresos, cada grupo social se moverá a lo largo de la curva particular que socialmente le corresponde.

En el agregado, la predicción del modelo es que si aumenta el nivel de capital humano de la sociedad sigma, el nivel de ingreso se elevará, pero la desigualdad en los ingresos no desaparecerá, aun si se lograra la igualdad en años de educación entre los grupos sociales. Igual predicción se obtiene para las dos sociedades omega y épsilon, aunque el equilibrio general será con un grado de desigualdad que no es tan pronunciada como en sigma.

UN MODELO DINÁMICO

Consideremos ahora un modelo de equilibrio general dinámico de cada sociedad, que incorpora la teoría del capital humano. Los años de escolaridad de los individuos será ahora una variable endógena. Esta decisión implica que el stock de capital humano de la economía se vuelve también endógeno.

Podemos fácilmente transformar el modelo estático presentado anteriormente en otro que sea dinámico. Todo lo que tenemos que hacer

es introducir el supuesto que existe una relación intertemporal entre algunas de las variables endógenas. Ahora, la cantidad demandada de años de educación de este periodo, así como la cantidad de capital humano acumulado, dependerán del nivel del ingreso nacional y de su distribución entre los grupos sociales, y determinarán el stock de capital humano del periodo siguiente.

Consideremos la solución de equilibrio general del modelo estático, señalado arriba, como la solución del periodo 1. Los valores de equilibrio de las variables endógenas se refieren a la producción y distribución, que incluye también el ingreso del gobierno —dada la tasa de impuestos—. Con estos valores conocidos, los individuos demandarán capital humano y el gobierno financiará la provisión de educación como bienes públicos. Así se determinará la cantidad de años de escolaridad que cada individuo desea y puede pagar, y se producirá la acumulación de capital humano, medido por un aumento en los años de educación de la gente.

La acumulación de capital humano que implica esa educación se agregará al stock inicial de capital humano y un nuevo stock estará disponible para la producción en el periodo 2. Con este nuevo stock de capital humano habrá un nuevo equilibrio general en la producción y distribución en el periodo 2, que a su vez determinará las cantidades demandadas y ofrecidas de años de escolaridad. La acumulación de capital humano que esta educación implica se agregará al stock anterior para la producción de bienes en el periodo 3, y así sucesivamente.

Desde una condición inicial, y dados los valores de las variables exógenas, la economía mostrará una trayectoria del ingreso nacional y su distribución. Esta trayectoria ocurre por el solo paso del tiempo, pues las variables exógenas se mantienen fijas. Esta trayectoria se denomina el equilibrio dinámico. Nótese que bajo nuestros supuestos, el equilibrio dinámico no es sino una secuencia de situaciones de equilibrio estático.

Ciertamente, en este proceso individual de acumulación de capital humano, los periodos deben entenderse como generaciones. Es un

proceso de larga duración. Las acciones de esta generación tendrán efectos recién en la siguiente. Y las acciones de esta segunda generación tendrán efectos recién en la tercera generación, y así sucesivamente. La acumulación que hace una familia en capital humano toma tiempo, mucho más tiempo que el que le toma a una firma acumular capital físico. Mientras una firma puede tener una nueva máquina instalada en una semana, una familia tendrá un ingeniero electrónico recién en veinte años. En el agregado, sin embargo, la sociedad acumula capital humano todos los años.

Al igual que en el modelo estático, en el modelo dinámico cada grupo social se moverá a lo largo de su curva particular, la que relaciona años de educación con ingresos. La razón es simple: la trayectoria que muestra el equilibrio dinámico no es sino una secuencia de situaciones de equilibrio estático. No existe ningún mecanismo en la sociedad que vaya a juntar endógenamente las diferentes curvas de cada grupo social en uno solo para toda la sociedad.

La predicción empírica del modelo dinámico es que las trayectorias de ingresos de los grupos sociales serán curvas crecientes en el tiempo, pero serán curvas paralelas, sin una tendencia a la convergencia entre ellas. En la sociedad sigma, las trayectorias de los tres grupos sociales mostrarán que los ingresos promedios crecen a través del tiempo, pero el orden escalonado de las curvas se mantendrá (A-Y-Z) sin que logren juntarse. Amigo lector: construya su propio gráfico, midiendo el tiempo en el eje horizontal y los ingresos en el eje vertical; trace tres curvas de pendiente positiva y paralelas, denominadas A, Y, Z. ¿Qué le dicen con respecto a la desigualdad entre los grupos sociales a través del tiempo?, ¿qué le dicen con respecto a la pobreza? Para el caso de las sociedades épsilon y omega solo habrá un sistema de dos curvas, pero con la misma característica.

El nuevo modelo de equilibrio general de la teoría sigma, que incorpora la teoría del capital humano, predice que el proceso educativo no es igualador de capital humano y que el mecanismo del mercado no es igualador de ingresos para el mismo capital humano. En suma,

introduciendo capital humano en el proceso económico, los modelos no predicen que el grado de desigualdad tenga una tendencia a reducirse de manera endógena. La misma conclusión se aplica a las teorías omega y épsilon. Los modelos predicen falta de movilidad social. La posición relativa de los ingresos, y de la posición social, tiende a transmitirse de padres a hijos.

A diferencia de lo que ocurre en el proceso biológico, donde las tallas de los padres se transmiten a los hijos con una tendencia hacia la igualdad, es decir, donde en cada generación sucesiva las diferencias se van reduciendo, no parece existir tal tendencia en los ingresos de los grupos sociales en el proceso económico. El biólogo inglés Francis Galton llamó a este proceso que conduce a la población hacia una talla más hemogénea la línea de regresión. Los modelos dinámicos de la teoría del capital humano no predicen esa línea de regresión.[7]

Falsación empírica

Los modelos estáticos con capital humano predicen que una reducción en los años de educación no conducirá a una reducción significativa en el grado de desigualdad en la sociedad. Según las estadísticas del Banco Mundial, en América Latina las tasas de matrícula han aumentado de manera espectacular, tanto en la primaria como en la secundaria, por varias décadas. Sin embargo, una regularidad es que la desigualdad en los ingresos se mantiene casi constante; el coeficiente Gini se ha mantenido alrededor de 0,50 en ese periodo, como mostramos en el segundo capítulo. Este dato de la realidad es consistente con la predicción del modelo para el caso de la teoría sigma.

Los mecanismos de exclusión que operan en el sistema educativo y en los mercados básicos, y que están a la base del poco efecto igualador del

[7] El término «regresión», que se usa ahora en estadística y econometría, pero con el significado de una tendencia de las observaciones hacia la media, viene de Galton.

sistema educativo, no han sido muy estudiados. Pero los pocos estudios que existen corroboran la predicción del modelo teórico presentado aquí. Así, un estudio para el Perú encontró que las relaciones entre el ingreso y los años de educación es positiva, pero las relaciones son diferentes por grupos sociales, en el orden que predice el modelo; encontró además que la población indígena —la población Z— se encuentra todavía excluida de la educación, pues la gran mayoría no ha logrado pasar el nivel secundario, cuando la clase alta y media alta ya ha llegado casi al techo de años de educación (Figueroa 2006). Otro estudio para Brasil encontró relaciones similares, utilizando como población Z a la población negra (Telles 1993).

Es común encontrar en la literatura económica que las diferencias salariales entre grupos raciales se atribuyen a la hipótesis —sin teoría— de la discriminación salarial. La evidencia que le sirve de sustento empírico es encontrar que en el mercado laboral se pagan diferentes salarios según la raza del trabajador, para una misma cantidad de años de educación.

El modelo presentado aquí invalida esa hipótesis, pues muestra que educación no es lo mismo que capital humano. Las firmas en su búsqueda de la ganancia tenderán a pagar igual salario por igual capital humano, no por igual cantidad de años de educación. Incluso, estos modelos sugieren que la diferencia de ingresos entre grupos étnicos, entre blancos e indígenas, no se debe tanto a que los ingenieros blancos ganen más salarios que los ingenieros indígenas, sino a que el capital humano de los primeros es superior a los segundos, y a que la proporción de ingenieros en la población blanca es mayor que en la población indígena.

Los modelos dinámicos con capital humano predicen falta de movilidad social. Los estudios empíricos sobre movilidad social han mostrado que, en efecto, la movilidad intergeneracional es muy débil, que en general los hijos tienden a heredar la posición relativa de sus padres. Así, varios estudios empíricos sobre los Estados Unidos han encontrado que las correlaciones entre los ingresos de los padres y de los hijos son positivas (Solon 1992 y Zimmerman 1992).

Conclusiones

En este capítulo se ha desarrollado una teoría del capital humano. Esta teoría intenta explicar las causas y consecuencias de la acumulación de dicho capital. A nivel del análisis microeconómico, un modelo de esta teoría ha podido explicar por qué existe desigualdad en la distribución del capital humano entre grupos sociales. Ha mostrado que el proceso educativo no es igualador en el capital humano. También ha mostrado que la desigualdad en los ingresos magnifica la desigualdad en el capital humano; es decir, el mecanismo del mercado no es igualador de ingresos.

Al nivel del análisis macroeconómico, del equilibrio general, la teoría del capital humano se ha incorporado a los modelos estáticos y dinámicos para cada sociedad épsilon, omega y sigma. Los modelos estáticos predicen que mayores niveles de educación llevan a mayores niveles de ingreso, pero que una reducción en las diferencias en años de educación no implican una reducción en la desigualdad en la distribución de ingresos. También predicen que aun si se lograra la igualdad en años de educación de manera exógena, la desigualdad en ingresos no desaparecería. Los modelos dinámicos han mostrado un resultado similar. No existe una tendencia a la reducción de la desigualdad de manera endógena.

Los datos existentes de la realidad tienden a corroborar esas predicciones. Aunque hay que notar que la literatura empírica internacional sobre este tema es muy escasa. Los datos de la realidad no han refutado las predicciones del modelo de la teoría del capital humano; por lo tanto, no existe razón alguna para rechazar esta teoría en esta etapa de la investigación.

Una razón de estos resultados es que cada grupo social avanza en curvas o trayectorias distintas y desiguales que relacionan educación a ingresos, tal como se ha mostrado en la figura 7.2. Otro sería el resultado si existiera una única curva que relacionara educación con ingresos para toda la sociedad. La igualdad en años de educación implicaría

igualdad en ingresos. En este caso se podría decir que existe igualdad de oportunidades en la sociedad. Luego lo único que necesitarían los pobres para igualar a los ricos sería acumular más años de educación. Este es implícitamente el paradigma en el discurso de las elites económicas y políticas. Los modelos presentados aquí tienen predicciones diferentes que son corroboradas por los datos: incluso si se lograran igualar los años de educación, no habría igualdad de ingresos. Incluso, los modelos predicen que no hay mecanismo que reduzca endógenamente las distintas curvas a una sola. No existe, por lo tanto, igualdad de oportunidades.

Estas predicciones son válidas para todo tipo de sociedad capitalista. Ciertamente, las relaciones cuantitativas entre educación e ingresos serán diferentes en diferentes sociedades, pero la naturaleza de la relación es la misma. Dada las condiciones iniciales —dotación de factores productivos y la desigualdad inicial—, en el proceso de acumulación del capital, la democracia y el mercado, las dos instituciones básicas del capitalismo, no conducen a la igualdad en la sociedad. El sistema democrático que opera a través del proceso educativo no es igualador, así como tampoco lo es el sistema de mercado.

En el capítulo siguiente se analizará el proceso de desarrollo económico, donde tanto el capital físico como el capital humano son endógenos. Esto se hará para cada sociedad y luego para el conjunto del capitalismo. Así llegaremos a una teoría unificada del capitalismo.

Capítulo 8

Teoría unificada del capitalismo

Este capítulo explicará el desarrollo económico del capitalismo. El concepto de desarrollo económico que utilizaremos deberá guardar consistencia con el campo de la ciencia económica; más precisamente, deberá tomar en cuenta las dos variables endógenas fundamentales del proceso económico: la producción y la distribución. Para fines operativos, producción y distribución se pueden transformar en nivel del producto por persona y grado de desigualdad en la distribución del producto total. Para una sociedad dada, entonces, cuanto mayor es el nivel del producto por persona y menor el grado de desigualdad, mayor será su nivel de desarrollo económico. Entre sociedades, estas dos variables nos indicarán el orden que ocupan en el nivel de desarrollo económico.

Las regularidades empíricas 6 y 7, que establecimos en el segundo capítulo, se refieren al capitalismo en su conjunto. Se refieren a la existencia y persistencia en las diferencias en los niveles de ingresos y en el grado de igualdad entre el primer mundo y el tercer mundo. En ese sentido, se puede señalar que una de las regularidades empíricas del capitalismo es la existencia y persistencia en las diferencias del nivel de desarrollo entre el primer mundo y el tercer mundo.

La división del mundo capitalista entre los países más desarrollados y los menos desarrollados es el fenómeno social de nuestro tiempo. Y es el fenómeno que nos falta explicar. ¿Cuáles son los factores —las variables exógenas— que explican las diferencias en el desarrollo económico

de las naciones? Naturalmente, la explicación implica tener una teoría unificada, esto es, una teoría del sistema capitalista considerado como un todo.

Naturaleza de una teoría unificada

Hasta ahora hemos desarrollado tres teorías que explican el funcionamiento de cada uno de los tres tipos de sociedades en que hemos dividido el mundo capitalista. Estas tres sociedades abstractas que hemos construido —épsilon, omega y sigma— son, por lo tanto, teorías parciales del capitalismo. ¿Será posible construir una teoría unificada sobre el sistema capitalista, considerado como un todo, sobre la base de estas teorías parciales?

¿Cómo se construye una teoría unificada? Según el *principio de Moore,* establecido por el matemático estadounidense del mismo nombre, si los supuestos primarios de las teorías parciales son lógicamente consistentes entre sí, estas teorías darán lugar a una teoría unificada. El conjunto de supuestos primarios de la teoría unificada estará constituido por el núcleo de los supuestos primarios de las teorías parciales.

Todo lo que queda por mostrar, entonces, es que las proposiciones alfa de cada teoría parcial dan lugar a un núcleo de proposiciones alfa comunes que son lógicamente consistentes entre sí. Revisando las proposiciones alfa de cada teoría, que aparecen en el cuarto capítulo, podemos ver que, en efecto, este núcleo existe. Tenemos entonces una teoría unificada.

La teoría unificada intentará explicar el sistema capitalista como un todo. La teoría unificada tendrá que ser, como toda teoría, refutable. Y para ello se tendrán que construir modelos de la teoría unificada. Estos modelos producirán predicciones empíricas que luego tendrán que ser contrastadas con las regularidades empíricas, y así evaluar su validez. Si la teoría fuese, en efecto, refutada por los datos de la realidad, tendríamos un problema de agregación, pues tendríamos teorías parciales que

explican sociedades particulares del sistema capitalista, pero no tendríamos una teoría unificada que explique el conjunto.

¿Sería esto posible? ¿Sería posible que buenas teorías parciales no conduzcan a una buena teoría unificada?

Este problema constituye el problema fundamental en la física actual. Existen dos buenas teorías parciales, una que explica el mundo de los cuerpos grandes —la teoría de la relatividad— y otra que explica el mundo de los cuerpos pequeños, subatómicos —la teoría cuántica—; sin embargo, ambas teorías no son consistentes entre sí, es decir, ambas no pueden ser verdaderas. El trabajo de la física teórica actualmente consiste en encontrar una teoría unificada, que se denomina la teoría del todo.

La tarea que vamos a acometer en este capítulo es mostrar que, en efecto, tenemos una teoría del todo del capitalismo. En la construcción de una teoría unificada existen entonces dos pasos. El primero es generar una teoría unificada de varias teorías parciales. El segundo consiste en construir un modelo y someter sus predicciones al proceso de falsación. Comencemos por el primero paso.

Proposiciones alfa de la teoría unificada

Las proposiciones alfa de cada una de las tres teorías parciales —épsilon, omega y sigma— fueron presentadas en el cuarto capítulo. A partir de ellas, se puede construir un núcleo de proposiciones consistentes entre sí, que constituirán las proposiciones alfa de la teoría unificada, los supuestos primarios sobre una sociedad capitalista abstracta. Ellas son:

> **Contexto institucional.** (a) Reglas: los individuos que participan en el proceso económico están dotados de activos económicos y sociales; los activos económicos están sujetos a los derechos de la propiedad privada; la gente intercambia bienes sujetos a las reglas del mercado, que incluye la norma de que los salarios nominales no pueden bajar. (b) Organizaciones: familias, firmas y el gobierno.

Condiciones iniciales. Los individuos están dotados con cantidades desiguales de activos económicos y sociales. Existen dos clases sociales: capitalistas y trabajadores.

Racionalidad de los agentes. Los individuos actúan guiados por la motivación del interés propio.

Tolerancia social a la desigualdad. Los individuos tienen una tolerancia limitada a la desigualdad. Cuando la desigualdad supera sus umbrales de tolerancia, los individuos reaccionan y buscan reducir la excesiva desigualdad.

Este conjunto de proposiciones alfa solo incluye el núcleo de las teorías parciales. Existen proposiciones alfa de las teorías parciales que no aparecen en el núcleo porque son particulares a cada sociedad. Podemos reconstruir una sociedad abstracta particular —épsilon, omega o sigma— agregando a este núcleo sus proposiciones alfa particulares. Así, las sociedades abstractas parciales se constituyen en teorías parciales o particulares de la teoría unificada. Los supuestos del núcleo son los que caracterizan al sistema capitalista abstracto.

Los supuestos primarios de la teoría unificada del capitalismo están constituidos por el conjunto de las proposiciones alfa listadas anteriormente. Para hacerla refutable, necesitamos construir un modelo de la teoría unificada. Como todo modelo, se debe formar agregando a los supuestos primarios un conjunto de supuestos auxiliares que sean consistentes. Este modelo de la teoría unificada dará lugar a predicciones empíricas refutables, las cuales serán sometidas a la refutación empírica, cuyo resultado nos dirá si se puede aceptar el modelo, y por consiguiente, la teoría. El segundo paso a realizar en la construcción de la teoría unificada es la construcción de modelos que nos permitan explicar las regularidades 6 y 7. Dos modelos explicativos, uno estático y otro dinámico, se presentan a continuación.

Un modelo estático de la teoría unificada

En un sentido estático, la sexta regularidad dice que el ingreso por persona en el primer mundo es superior al del tercer mundo, y que cambios en las variables exógenas no van a eliminar o reducir significativamente esta diferencia; por su parte, la sétima regularidad dice que el grado de desigualdad en el primer mundo es menor que en el tercer mundo y que cambios en las variables exógenas no van a eliminar o reducir de manera significativa esta diferencia.

Se puede construir un modelo estático de la teoría unificada integrando los modelos estáticos de las teorías parciales presentados en los capítulos anteriores. Las variables endógenas del modelo son los niveles de ingreso medio y los grados de desigualdad. Las variables exógenas, comunes a las tres sociedades abstractas, son las que se refieren a sus condiciones iniciales: la dotación inicial de factores productivos y el grado de desigualdad inicial. Los efectos de cambios en estas variables exógenas sobre las endógenas las hemos analizado para cada sociedad abstracta de manera separada. La estática comparativa opera para cada modelo parcial, pero no para el agregado del sistema. No tenemos forma de hacer variar las variables exógenas para todo el sistema y analizar su efecto en cada economía abstracta.

Sin embargo, podemos analizar el sistema capitalista a partir de los tres modelos estáticos parciales tomados en conjunto y mostrar si pueden predecir el orden que establecen las regularidades empíricas. Para ello, introducimos supuestos auxiliares. Primero, supondremos que las tres sociedades producen el mismo producto, el bien B. Supondremos, luego, que la tecnología para producir el bien B en las firmas capitalistas es la misma en las tres sociedades. Es decir, las mismas cantidades de factores producen las mismas cantidades de producto total en el sector capitalista de cada una de las tres sociedades abstractas. Además, la tecnología es tal que el doble de cantidad de factores produce el doble de producto total —supuesto de rendimientos a escala constantes—, lo cual implica que el producto por trabajador depende

solo de la cantidad de capital por trabajador; y, finalmente, que al doble de cantidades de capital por trabajador le corresponde mayores cantidades de producto por trabajador, pero no el doble, sino menos que este —principio de los rendimientos decrecientes del factor—.

Una implicancia de este supuesto sobre la tecnología es que el producto por trabajador depende positivamente de la cantidad de capital por trabajador de la economía. A la misma relación capital-trabajo le corresponderá el mismo nivel de producto por trabajador. El supuesto de que el doble de capital y de trabajo dobla el producto total lleva a este resultado.

Las condiciones iniciales en cuanto a las dotaciones de factores productivos de las tres sociedades abstractas son tales que la cantidad de capital por trabajador —donde capital incluye capital físico y capital humano— en épsilon es mayor que en omega, y en omega mayor que en sigma. Dado los supuestos sobre la tecnología, obtenemos la siguiente predicción del modelo conjunto: el ingreso por trabajador sigue el mismo orden que el del capital por trabajador; es decir, el nivel de ingreso medio en épsilon es mayor que en omega, y mayor en omega que en sigma.

Esta relación se obtiene tomando en cuenta solo al sector capitalista de cada sociedad. Pero sabemos que, tanto en omega como en sigma, donde existen sectores de subsistencia, el producto por trabajador de la economía es inferior al producto por trabajador de su sector capitalista. Por lo tanto, la predicción que se estableció en base a los sectores capitalistas solamente, se verá ahora reforzada; es decir, entre las economías capitalistas, a mayor cantidad de capital por trabajador le corresponderá un mayor nivel de producto por trabajador. El ingreso por trabajador también sigue el mismo orden que el ingreso por persona, pues a mayor población habrá mayor cantidad de trabajadores.

Tenemos así una proposición beta, pues en principio puede ser falsa. Dado que empíricamente el primer mundo tiene una mayor dotación de capital por trabajador que el tercer mundo, llegamos ahora a reconstruir la sexta regularidad del capitalismo: el primer mundo es más rico que el tercer mundo.

La sexta regularidad empírica, establecida en el segundo capítulo, se refiere al primer y tercer mundo. Allí no se hizo la distinción entre las dos secciones del tercer mundo que viene de la teoría: países sin legado colonial importante —sociedad tipo omega— y con legado colonial —sociedad tipo sigma—. Ahora podemos dar la visión de conjunto. Los países sin legado colonial importante son pocos: Argentina, Corea del Sur, Israel, Tailandia, Taiwán y Uruguay. En 1999, estos países tenían un nivel de ingreso promedio de cerca de 7.000 dólares, mientras que esta cifra en el resto del tercer mundo, con legado colonial, era de 1.200 y en el primer mundo era de 25.700 (Banco Mundial 2001: cuadro 1).

Aunque la diferencia de ingresos entre el primer mundo es resultado de nuestra definición —hemos llamado primer mundo al grupo de países más ricos—, lo que es refutable es si ese orden está asociado a las dotaciones factoriales de los países, como predice el modelo estático de la teoría unificada. Los datos son, en efecto, consistentes con la predicción del modelo.

Sobre el grado de desigualdad, el modelo también predice un cierto orden entre las sociedades abstractas. En cada sociedad, el grado de desigualdad en la distribución de los flujos del ingreso nacional depende positivamente del grado de desigualdad inicial en la distribución de los activos entre los individuos. A mayor desigualdad inicial en activos, le corresponderá mayor desigualdad en los flujos de ingresos.

Dado los supuestos sobre las condiciones iniciales de cada sociedad abstracta, la desigualdad inicial en sigma es claramente mayor que en épsilon. Omega es una sociedad homogénea —a semejanza de épsilon—, pero en su dotación factorial inicial muestra superpoblación —a diferencia de épsilon—, que implica mayor desigualdad en la distribución del stock de capital entre individuos; entonces, se puede decir que la desigualdad inicial en omega es mayor que la de épsilon. Con relación a sigma, omega es también superpoblada pero mucho menos que sigma, por lo cual su desigualdad inicial sería menor; por otro lado, dado que omega es socialmente homogénea y sigma es socialmente

heterogénea, su desigualdad inicial también sería menor, y reforzaría el efecto anterior. Por lo tanto, el orden de la desigualdad inicial es así: sigma es la más desigual, épsilon la menos desigual y omega se encuentra entre medio. La mayor diferencia en la desigualdad se encuentra entre sigma y épsilon.

Al igual que en el caso de la sexta regularidad, en la sétima se necesita añadir las diferencias en desigualdad entre las dos secciones en que se ha dividido el tercer mundo para comparar los datos con las predicciones. Para el grupo de países sin legado colonial importante, listado arriba, existe información solo para Corea del Sur, Tailandia y Taiwán, con una media en el coeficiente de Gini de 0.36 para el grado de desigualdad en la distribución del ingreso nacional. El valor de este coeficiente en el tercer mundo (con legado colonial) es de 0.45 y de 0.33 en el primer mundo (Deninger y Squire 1996: cuadro 1).

Debido a que el orden en el grado de desigualdad inicial en la distribución de los activos es tercer mundo - primer mundo, se corrobora que la desigualdad en la distribución de los flujos de ingresos sigue el mismo orden, tal como predice el modelo estático de la teoría unificada. Como se puede ver, el valor de los países tipo omega cae entre los dos valores, aunque más cercano al del primer mundo. Los datos de la realidad son entonces consistentes con la predicción del modelo.

Un comentario adicional. Las dos relaciones empíricas que hemos derivado del modelo estático de la teoría unificada se refieren a situaciones de equilibrio de largo plazo en cada tipo de sociedad capitalista. Tal como se mostró en el cuarto capítulo, en cada sociedad, dada su dotación factorial —una variable exógena—, el ingreso medio de equilibrio de largo plazo queda determinado; cambios en las otras variables exógenas, generan solo variaciones alrededor de este valor. Asimismo, dada su desigualdad inicial —una variable exógena—, el grado de desigualdad en la distribución del flujo de ingresos de largo plazo queda determinado y cambios en las otras variables exógenas generan solo variaciones alrededor de este valor.

En suma, el modelo estático de la teoría unificada, vista como una integración de los modelos estáticos de las tres sociedades capitalistas que hemos presentado en el cuarto capítulo, predice el orden que se observa entre los grupos de países, tanto en el nivel de ingreso promedio como en el grado de desigualdad. Los datos de la economía capitalista tomada como un todo —las regularidades 6 y 7— no refutan las predicciones del modelo estático. Los países del primer mundo tienen un mayor nivel de ingreso que los países del tercer mundo porque sus trabajadores están, en promedio, equipados con mayor capital físico y humano que los trabajadores del tercer mundo; asimismo, la desigualdad en los ingresos es menos pronunciada en los países del primer mundo porque su distribución de activos económicos y políticos no es tan desigual comparada a los países del tercer mundo.

Un modelo dinámico de la teoría unificada

Las regularidades 6 y 7 tienen otro componente que es dinámico. Ahora se trata de explicar la persistencia de las diferencias observadas tanto en el nivel del ingreso medio como en el grado de desigualdad. ¿Por qué *persisten las diferencias entre las dotaciones factoriales* y las desigualdades iniciales? Para tal efecto construiremos un modelo dinámico de la teoría unificada.

Consideremos, como en el modelo estático, que el sistema capitalista está compuesto de tres sociedades abstractas: épsilon, omega y sigma. La tecnología de producción del bien B es la misma para las tres sociedades y tiene dos características: los rendimientos a escala son constantes y los rendimientos del factor son decrecientes.

Para el modelo dinámico, haremos los siguientes supuestos auxiliares sobre el proceso de acumulación de capital. En cada sociedad abstracta, la nueva tecnología está incorporada en la acumulación del capital físico, lo que hace que la adopción del cambio tecnológico sea también endógena; además, este proceso de acumulación del capital

y de cambio tecnológico ocurre solo en el sector capitalista de cada economía. Como parte del contexto mundial, supondremos que la tasa de progreso tecnológico es exógena y que existe perfecta movilidad del capital financiero entre economías.

Las variables exógenas del modelo serán de dos tipos. Las que se refieren a cada sociedad, incluirán sus condiciones iniciales: dotación inicial de factores productivos y desigualdad inicial; y las que se refieren al sistema en su conjunto, incluirán la tasa de crecimiento de la tecnología mundial y también el nivel mundial de inversiones.

La sociedad épsilon

Veamos primero el proceso de desarrollo en la sociedad épsilon. Entre las condiciones iniciales supondremos que la economía se encuentra con pleno empleo efectivo en el mercado laboral. Dada las condiciones iniciales, el equilibrio general del primer periodo determina la cantidad del producto total y su distribución; también la economía invierte en capital físico y capital humano en este periodo. La economía inicia el segundo periodo con un mayor stock de capital y determina la producción y distribución, que a su vez determina la inversión en capital físico y humano, que lleva a la economía a iniciar el tercer periodo con un mayor stock de capital, y así sucesivamente. Debido a la acumulación de capital, que incorpora la nueva tecnología, el producto total crece de manera endógena periodo tras periodo. Existe un equilibrio dinámico que es una secuencia de situaciones de equilibrio estático. En esencia, este es el proceso de crecimiento endógeno del producto total de la economía épsilon.

Ahora introducimos algunas precisiones. En el proceso de crecimiento endógeno del producto total también aumenta la población. La pregunta ahora es si la población crece al mismo ritmo que el producto, pues la pregunta del crecimiento económico no es si el producto total crece a través del tiempo y de manera endógena, sino si el producto por persona o por trabajador crece de manera endógena. Dada una tasa de

crecimiento de la población —y de la oferta de trabajadores—, la tasa de crecimiento del producto por trabajador será igual a la diferencia entre la tasa de crecimiento del producto total menos la tasa de crecimiento de la población. Si el producto total crece a la tasa de 5% anual y la población crece a la tasa de 3% anual, entonces el producto por persona crecerá al 2% anual. Como suponemos que la oferta de trabajadores crece a la misma tasa de la población, el producto por trabajador también crecerá al 2% anual. Llamaremos *crecimiento económico* al aumento sostenido en el producto por trabajador que experimenta la sociedad.

Si los trabajadores y el stock de capital crecieran al 2% anual, el producto total también crecería al 2%, debido a los rendimientos a escala constantes. Pero entonces el producto por trabajador estaría fijo; no habría crecimiento. ¿De dónde proviene el crecimiento económico? Suponga que el progreso tecnológico creciera al 3%, entonces el producto total crecería, digamos, al 5% anual. El producto por trabajador, es decir la productividad laboral, crecería al 3%. La respuesta es, entonces, que el crecimiento se origina en el progreso tecnológico. Si este progreso fuese nulo, no habría crecimiento económico. Supondremos que el trabajador se vuelve más productivo con el progreso tecnológico.

Dada la tasa de inversión —es decir, la cantidad de capital por unidad del producto total que se agrega al proceso productivo cada año—, habrá un equilibrio dinámico del producto por trabajador. La condición de equilibrio es que el stock de capital debe crecer a la misma tasa que resulta de sumar las tasas de crecimiento de la población y del progreso tecnológico. Utilizando los datos del ejemplo anterior, el capital debe crecer al 5% anual; por lo tanto, la relación capital-trabajo crecerá a una tasa de 3% anual.

Si la dotación inicial de factores satisface la condición de equilibrio en el periodo inicial, el producto por trabajador crecerá a la misma tasa en los periodos siguientes (al 3% anual en el ejemplo). Esta trayectoria dará lugar a la *frontera de crecimiento* del producto por trabajador. Esta frontera es una curva que muestra la trayectoria del producto por trabajador a través del tiempo. Es una frontera porque la sociedad no

puede estar por encima de esa trayectoria, aunque puede estar por debajo temporalmente debido a que sus condiciones iniciales no fueron las apropiadas.

Amigo lector, construya la frontera de crecimiento en un gráfico, donde el eje vertical mide producto por trabajador y el eje horizontal el tiempo en años. Trace una curva creciente que se origine en un punto del eje vertical. De este mismo punto trace una curva lineal y comprare con la anterior. ¿En cuál de ellas la tasa de crecimiento anual puede ser constante?

Como veremos más adelante, un factor que determina la frontera de crecimiento es la tasa de inversión. Una economía tendrá una frontera superior si invierte en capital una magnitud que es equivalente al 20% del producto total, que si fuese 10%. Debería quedar claro que estos porcentajes no se refieren a tasas de ahorro nacional sino a la inversión. En una economía donde existe movilidad del capital, el factor relevante es la inversión, pues la inversión es independiente del ahorro nacional; en una economía sin movilidad de capitales, en cambio, el factor relevante es el ahorro nacional. También debería quedar claro que aquí el término «capital» incluye capital físico y capital humano, de manera que «inversión» significa el gasto en ambos tipos de capital.

¿Qué factores determinan el *nivel* —más arriba o más abajo— de la curva de la frontera de crecimiento? El nivel de la curva depende de los coeficientes de la economía que determinan el crecimiento de los factores productivos: la tasa de inversión, la tasa de crecimiento de la población y el nivel inicial de la tecnología. El equilibrio dinámico implica que todos los factores productivos deben crecer a la misma tasa, de modo que el producto por trabajador crece a la tasa del progreso tecnológico. Esta condición a su vez implica unas condiciones iniciales particulares, tal como una dotación factorial inicial que sea consistente con el equilibrio del periodo inicial.

La frontera de crecimiento es, entonces, una curva que muestra la trayectoria del producto por trabajador en el tiempo, que constituye el equilibrio dinámico de la economía: el producto por trabajador crece

por el puro paso del tiempo, manteniendo fijas las variables exógenas. Pero la economía de épsilon se moverá a lo largo de la frontera si y solo si tiene una dotación factorial inicial (capital por trabajador) que corresponde al equilibrio inicial de largo plazo. Si la dotación inicial de capital por trabajador está por debajo de este equilibrio, el producto por trabajador en el periodo inicial estará por debajo de la frontera.

Como el equilibrio dinámico es una secuencia de situaciones de equilibrio estático, y este equilibrio es estable, el equilibrio dinámico también será estable. Asimismo, si la relación capital por trabajador inicial —y el correspondiente ingreso medio— está por debajo de la de equilibrio, la economía tenderá a restaurar el equilibrio espontáneamente, aumentando esa relación. Desde su posición inicial, la economía se dirigirá hacia la frontera de crecimiento.

En ese periodo de transición hacia la frontera, la economía crecerá temporalmente a una tasa superior a la que crece el producto por trabajador a lo largo de la frontera. ¿Por qué existe este efecto? En el periodo de transición, la economía aumentará aun más la relación capital por trabajador; luego, a mayor capital por trabajador, habrá mayor producto por trabajador. Una vez en la frontera, la economía crecerá a lo largo de ella, a la tasa de crecimiento de la tecnología, que esta exógenamente determinada.

Este es el proceso de crecimiento endógeno de la sociedad épsilon. Recordemos que la economía de épsilon se compone de un sector capitalista solamente.

La sociedad omega

En la sociedad omega existen dos sectores, el capitalista y el de subsistencia. La frontera de crecimiento estará determinada solo por el comportamiento del sector capitalista, y en los términos mostrados en épsilon. Los valores de las variables exógenas que determinan el nivel de la curva serán distintos, pero la tasa a la que crece la curva será la misma que la de épsilon: la tasa de crecimiento de la tecnología.

Omega es una economía sobrepoblada. La frontera de crecimiento muestra, por definición, una economía capitalista sin sobrepoblación. La razón es que la frontera muestra el equilibrio de largo plazo, donde el stock de capital y la población crecen a la misma tasa, y donde había pleno empleo efectivo en el periodo inicial.

Las condiciones iniciales de omega no corresponden entonces a las condiciones particulares que establece el equilibrio de largo plazo que está implícito en la frontera de crecimiento. La cantidad de capital por trabajador en el sector capitalista estará por debajo del de equilibrio. Pero como el equilibrio es estable, esa relación aumentará de manera espontánea hasta llegar al valor de equilibrio. En ese periodo de transición, el sector capitalista crecerá temporalmente a tasas superiores que la de la frontera. El capital crecerá a una tasa superior a la del equilibrio porque el empleo en el sector capitalista tiene que absorber no solo la nueva población sino también el exceso de oferta laboral existente. Desde sus condiciones iniciales, la economía de omega llegará a la frontera de crecimiento y una vez allí crecerá a la tasa que dicta el crecimiento de la tecnología.

En el periodo de transición, el sector capitalista irá eliminando el autoempleo que existe en el sector de subsistencia. Al final del periodo de transición, cuando la economía llega a la frontera del crecimiento, el sector de subsistencia habrá desaparecido. Y, así, la sociedad omega deviene en una sociedad épsilon de manera endógena. En el proceso de crecimiento económico, la sociedad omega no solo experimenta cambios cuantitativos, como un mayor nivel de ingreso medio, sino también cambios cualitativos.

La sociedad sigma

La sociedad sigma se compone de tres sectores, un sector capitalista, un sector de subsistencia Y, y un sector de subsistencia Z. También se puede decir que la sociedad sigma se compone de dos sectores: un sector omega —compuesto por el sector capitalista y el sector de

subsistencia Y— y un sector de subsistencia Z. Esta segunda definición es más útil para el análisis, pues ya sabemos el proceso de crecimiento en la economía de omega. En particular sabemos que la frontera de crecimiento corresponde a la frontera del sector capitalista, pues el sector de subsistencia Y desaparecerá en el periodo de transición.

En la sociedad sigma tal proceso no podría ocurrir. El sector de subsistencia Z no puede ser absorbido por el sector capitalista, pues el proceso de acumulación de capital humano deja siempre retrasado a los trabajadores Z, tal como se mostró en el capítulo 7. En el proceso de crecimiento económico, la sociedad sigma no devendrá en sociedad épsilon, pues los trabajadores Z no podrán ser transformados en trabajadores Y de manera endógena. En la sociedad sigma, el sector de subsistencia Y desaparecerá en el proceso de crecimiento económico, pero no el sector de subsistencia Z.

Dada la ausencia de interrelaciones entre el sector de subsistencia Z con el resto de la economía, sigma puede en parte operar como omega. El equilibrio dinámico en sigma será, entonces, la combinación de un sector capitalista —que opera como en omega— con un sector de subsistencia Z. La sociedad sigma experimentará cambios cuantitativos, como aumento en el producto por trabajador, modernización tecnológica, pero no sufrirá cambios cualitativos, seguirá siendo una sociedad sigma.

El producto por trabajador o ingreso medio en sigma es un promedio ponderado de los ingresos medios de los tres sectores. Existe un orden en estos valores, el ingreso medio del sector capitalista es mayor que el del sector de subsistencia Y, y este es mayor que el del sector de subsistencia Z. El ingreso medio de sigma se ubicará entre estos valores. La tasa de crecimiento del ingreso medio también es igual al promedio (ponderado) de las tasas de crecimiento de los sectores.

Una vez que el proceso de crecimiento en sigma conduce a la desaparición del sector de subsistencia Y, el ingreso medio será igual al promedio ponderado de los ingresos medios de los dos sectores solamente: capitalista y sector de subsistencia Z. De igual manera, la tasa

de crecimiento del ingreso medio será igual al promedio ponderado de las tasas de los dos sectores.

En sigma existirán dos fronteras de crecimiento, una del sector capitalista —con las características de la frontera que vimos en omega— y otra frontera para el conjunto de la economía de sigma. Los dos sectores capitalista y de subsistencia Y se moverán como si fueran una economía de omega: a partir de un punto inicial llegarán a la frontera capitalista, a partir del cual crecerán a la tasa dictada por la tasa de progreso tecnológico.

Una vez que se ha llegado a la frontera capitalista, el ingreso medio del sector capitalista será mayor que el del sector de subsistencia Z; por lo tanto, el ingreso medio ponderado de ambos será el ingreso medio de la sociedad sigma, que es un punto que pertenece ahora a la frontera de la sociedad sigma, obviamente por debajo de la frontera del sector capitalista. Si la frontera de sigma pudiera crecer a la tasa dictada por la tasa del progreso tecnológico, las dos curvas de las fronteras serían paralelas. Pero el sector de subsistencia Z crecerá a una tasa inferior; por lo tanto, la tasa de crecimiento de la sociedad sigma será inferior a la de su sector capitalista. En suma, en comparación a la frontera de su sector capitalista, la frontera de la sociedad sigma se puede representar por una curva de más bajo nivel y de menor pendiente.

Sobre la persistencia de las diferencias en los niveles de ingreso

Hemos presentado hasta aquí el proceso de crecimiento económico para cada una de las sociedades que conforman el sistema capitalista. Ahora toca ponerlas juntas en un solo modelo dinámico de la teoría unificada del capitalismo.

Las variables exógenas, comunes a todas las sociedades, son cuatro: la tasa de inversión, la tasa de crecimiento demográfico, la proporción de la población Z en la población total y el nivel de la tecnología inicial.

Ciertamente, los valores que toman estas variables exógenas son diferentes entre sociedades, y son estas las que determinan las diferencias en sus fronteras de crecimiento.

Pero, ahora debemos preguntarnos sobre los *factores últimos* que explican las diferencias en el crecimiento económico de las sociedades. ¿Serán esas cuatro variables exógenas los factores explicativos últimos? Introduciendo los resultados que hemos obtenido a lo largo del libro, podemos mostrar que esas variables exógenas constituyen más bien los *factores próximos* en el proceso de crecimiento económico, pues se les puede transformar en variables endógenas. Así, la tasa de inversión dependerá de la desigualdad inicial y del nivel mundial de inversiones. Esta relación la obtuvimos en los modelos teóricos que desarrollamos en los capítulos 6 y 7 sobre la acumulación de capital, tanto físico como humano.

La tasa de crecimiento de la población dependerá negativamente de la desigualdad inicial, pues las familias pobres tenderán a tener un mayor número de hijos que las familias ricas. Esta relación se infiere de los resultados obtenidos en el sexto capítulo, donde los trabajadores no tienen activos económicos y tampoco pueden obtenerlos por las restricciones que impone el funcionamiento de los mercados de crédito y de seguros. Su protección social vendrá entonces de buscar un tamaño grande de familia nuclear y de familia extensa, en la forma de redes sociales. Aquí ignoraremos el papel del Estado.

La proporción de la población Z en la población total dependerá de la desigualdad inicial, pues es el tipo de episodio fundacional de la sociedad capitalista el que genera o no la existencia de esta población, como se mostró en el cuarto capítulo.

Finalmente, el nivel inicial de la tecnología que se utiliza en la sociedad no se considerará arbitrario, sino que será parte de sus condiciones iniciales; en particular, dependerá de su desigualdad inicial. Como mostramos en el sexto capítulo, la acumulación de capital físico depende de la desigualdad inicial de la sociedad y la tecnología está incorporada en el stock de capital físico.

Consolidando estas relaciones nos quedamos, al final, con dos variables exógenas: la desigualdad inicial y el nivel mundial de inversiones. Debido a que el nivel mundial de inversiones es uniforme para todas las economías, las diferencias en la frontera de crecimiento de las economías estarán en sus desigualdades iniciales solamente.

En suma, la desigualdad inicial es el factor último que explica las diferencias en el crecimiento económico de largo plazo de las distintas sociedades capitalistas. Sociedades que se iniciaron al desarrollo capitalista como sociedades muy desiguales tendrán una frontera de crecimiento que se encuentra a un nivel por debajo de la frontera de sociedades menos desiguales. Las fronteras de crecimiento de las sociedades épsilon y sigma estarán, entonces, una debajo de otra; la frontera de omega se ubicará en el medio.

La trayectoria de crecimiento que sigue una sociedad depende, como vimos anteriormente, de su equilibrio inicial y de su frontera de crecimiento, pues desde ese punto inicial se dirigirá a la frontera, la llamada trayectoria de transición, para luego moverse a lo largo de la frontera. Recordemos que una frontera de crecimiento constituye el equilibrio dinámico de la economía de largo plazo.

Una sociedad épsilon que tenga un ingreso inicial por debajo del equilibrio se dirigirá desde ese punto a su frontera y luego se moverá a lo largo de ella. Cuanto más lejos esté el ingreso inicial del que corresponde a la frontera, la velocidad a la que se acercará a ella será mayor, es decir, su tasa de crecimiento temporal en el periodo de transición será mayor que el de largo plazo.

Una sociedad omega, que tenga el mismo grado de desigualdad inicial que el de épsilon, tendrá como frontera la que corresponde a la sociedad épsilon. Y desde el punto inicial de ingreso que tuviera, se dirigirá a esa frontera. Debido a las diferencias en la dotación inicial de factores productivos, la cantidad de capital por trabajador, el ingreso inicial de una sociedad omega estará por debajo del de una sociedad épsilon; por lo tanto, la sociedad omega crecerá en el periodo de transición

a una tasa mayor que épsilon. Una vez en la frontera, ambas economías crecerán a la misma tasa, dada por el crecimiento de la tecnología.

Una sociedad sigma tiene una frontera más baja y un ingreso inicial más bajo comparado a épsilon y omega. Por lo tanto, desde su posición inicial se dirigirá a su frontera. Cuanto más lejos esté de su frontera, su tasa de crecimiento será mayor en el periodo de transición. Pero queda claro que, bajo sus condiciones iniciales, nunca llegará a la frontera de épsilon.

Este es el resultado más importante del modelo dinámico de la teoría unificada. El modelo predice que, en el largo plazo, las sociedades tipo épsilon tienden a igualarse entre sí en el nivel de ingresos. También predice que las sociedades tipo omega tienden a acercarse a la frontera de épsilon y así tienden a igualarse en el nivel de ingresos. El modelo también predice que las sociedades tipo sigma no llegarán a convertirse en una sociedad épsilon, ni llegarán a alcanzar su nivel de ingresos. En el largo plazo, sigma crecerá a lo largo de un sendero que es distinto al de épsilon y omega.

Estas relaciones aparecen graficadas en la figura 8.1. En el eje vertical medimos el ingreso anual por persona (y) —toneladas del bien B por persona— y en el eje horizontal el tiempo en años (t). La curva E^*F^* representa la trayectoria de la frontera de crecimiento de la sociedad épsilon y la curva S^*R^* la de la sociedad sigma. Cada curva tiene una pendiente creciente que indica que el ingreso por persona crece a través del tiempo a una tasa anual.

Cada curva que representa la frontera de crecimiento se parece a la curva de ahorros de una cantidad de dinero que se coloca en un banco, donde la tasa de interés que paga el banco hace aumentar el dinero de manera creciente, pues el interés ganado en el año se suma al monto inicial, y sobre la nueva cantidad se ganan intereses que luego se suman al monto anterior, y así sucesivamente. Suponiendo una tasa de interés fija, se obtiene una curva de ahorros creciente. Considere ahora que la curva creciente del gráfico no muestra ahorro sino el producto por

persona y que la tasa de interés es igual a la tasa de crecimiento del producto por persona.

Una sociedad épsilon que tiene el ingreso inicial en el nivel E crecerá a lo largo de la curva EA, donde A es el punto de llegada a su frontera de crecimiento E^*F^*, y desde allí crecerá a lo largo de su frontera. Una sociedad sigma que tiene el ingreso inicial en el nivel S crecerá a lo largo de la curva SC, donde C es el punto de llegada a su frontera de crecimiento S^*R^*, y desde allí crecerá a lo largo de su frontera. Las sociedades épsilon convergerán a su respectiva frontera de crecimiento desde sus ingresos iniciales; de manera similar, las sociedades sigma convergerán a su respectiva frontera de crecimiento desde sus ingresos iniciales. Pero las sociedades sigma no llegarán a la frontera de la sociedad épsilon. Habrá convergencia relativa, pues cada sociedad llegará a su respectiva frontera; pero no habrá convergencia absoluta, pues las sociedades épsilon y sigma convergen a distintas fronteras.

En este modelo particular, la sociedad omega no tiene su propia frontera, sino que su frontera de crecimiento es la de épsilon; luego una sociedad omega que tenga el ingreso inicial en el nivel M crecerá a lo largo de la curva MB, donde B es el punto de llegada a la frontera de épsilon E^*F^*. En este caso, la sociedad omega converge a épsilon. Existe convergencia absoluta entre estas dos sociedades. La convergencia es cualitativa y cuantitativa: la sociedad omega se transforma en sociedad épsilon y llega a tener el mismo producto por persona que épsilon. La figura 8.1 resume las predicciones del modelo dinámico de la teoría unificada sobre las diferencias en el crecimiento económico de las tres sociedades.

La figura 8.1 muestra la trayectoria de largo plazo del producto por persona de cada tipo de sociedad capitalista. Luego, cualquier variación de corto plazo de esta variable se dará alrededor de la trayectoria de largo plazo. Luego, el modelo predice que las variables exógenas de corto plazo en cada sociedad, como la política monetaria o el coeficiente de términos de intercambio internacionales, tendrán efectos sobre las variaciones del producto por persona alrededor de la frontera

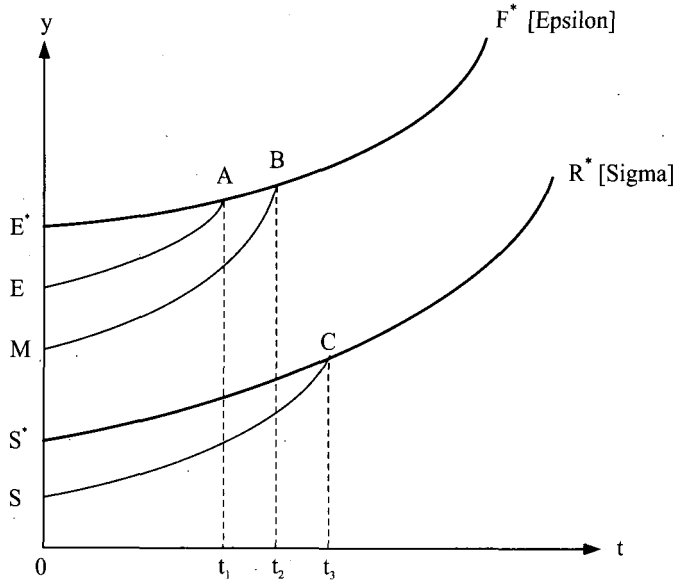

Figura 8.1 Trayectorias de crecimiento de las sociedades capitalistas

de crecimiento, pero no tendrán efectos importantes sobre la posición de la frontera misma. La posición de la frontera de crecimiento en las sociedades capitalistas depende de los factores estructurales de cada tipo de sociedad, como su desigualdad inicial.

¿Refutan los datos de la realidad las predicciones del modelo? Las teorías parciales mostraron que los países que forman el primer mundo se parecen a la sociedad épsilon y los del tercer mundo que tienen un legado colonial importante a la sociedad sigma. Las diferencias en los niveles de ingreso y su persistencia en el tiempo entre el primer mundo —sociedades épsilon— y el tercer mundo —sociedades sigma— están, en efecto, representadas en la figura 8.1. Estas relaciones son precisamente las que muestran los datos de la realidad, que aparece como la sexta regularidad empírica del capitalismo en el segundo capítulo.

El modelo dinámico no predice un orden en las tasas de crecimiento de los países en el tiempo. Esta tasa dependerá de su nivel de ingreso

inicial, de su distancia a su respectiva frontera de crecimiento y de los efectos de factores exógenos de corto plazo. Cualquier resultado estadístico sobre diferencias en las tasas observadas entre países no refutará el modelo. La predicción del modelo dinámico es sobre diferencias en niveles de ingresos entre países, sobre la desigualdad entre países en el largo plazo; no sobre las diferencias en tasas de crecimiento del ingreso en ciertos periodos de tiempo, como ha sido el interés de toda la literatura empírica reciente sobre el problema de la convergencia (por ejemplo, Barro y Sala-i-Martin 1995).

El modelo también predice que los países del tercer mundo sin un legado colonial importante —sociedades tipo omega— son los que pueden alcanzar el nivel de ingresos del primer mundo. Según la literatura internacional, historiadores económicos como Angus Madisson (1995), han mostrado que entre los países que lograron alcanzar a los países del primer mundo de hoy, desde un bajo nivel de ingreso, solo se puede mencionar a Japón. Es el único caso de convergencia absoluta. Japón, en efecto, tiene características de una sociedad omega, no tiene un legado histórico de dominación colonial; por el contrario, fue un país imperial. Japón se inició en el capitalismo como una sociedad relativamente igualitaria. El coeficiente de Gini promedio entre 1950 y 1995 es de 0.34, con una dispersión pequeña (Deninger y Squire 1996: cuadro 1). Según el modelo dinámico, Japón representa el caso de una sociedad omega que devino en épsilon.

Entre los países que definimos como omega en la actualidad, Taiwán y Corea del Sur son países que según Maddison están en ruta de llegar a tener el nivel de ingreso del primer mundo. Argentina y Uruguay son casos diferentes, pues pertenecían al grupo del primer mundo en una época y ahora se encuentran en el del tercer mundo. Según Maddison son los únicos casos de involución. En el resto de países de su muestra, los países que pertenecían al grupo de los ricos en 1820 siguen en ese grupo hoy día, y los países que pertenecían al grupo de los pobres siguen en esa misma posición actualmente.

Sobre la persistencia de las diferencias en desigualdades

¿Qué ocurre con la desigualdad en la distribución del ingreso nacional en el proceso de crecimiento económico?

Dado que la frontera de crecimiento constituye el equilibrio dinámico de la economía, habrá una distribución implícita a lo largo de la frontera, la cual se hará explícita ahora. En este modelo, producción y distribución se resuelven de manera simultánea en el proceso económico.

En la sociedad épsilon, la distribución del ingreso se refiere a la distribución del ingreso nacional entre ganancias y salarios. Una propiedad de la tecnología productiva que hemos supuesto —de rendimientos constantes a escala—, es que la productividad marginal es una proporción fija del producto promedio —o producto por trabajador—. Luego, ambos valores crecerán a la misma tasa, es decir, si el producto por trabajador crece al 3% anual, la productividad marginal también crecerá al 3% anual. Como en el equilibrio del mercado laboral, donde el salario real debe ser igual a la productividad marginal del trabajo, el salario real también crecerá al 3% anual.

La distribución de ingresos entre asalariados y capitalistas se mantendrá constante en el equilibrio dinámico. Usando las cifras del ejemplo anterior, utilizado en el caso de la sociedad épsilon, si el producto por trabajador y el salario real crecen a la misma tasa de 3% anual y la cantidad de trabajadores crece al 2%, la implicancia es que la masa salarial crecerá al 5%, igual a la del producto total y, como consecuencia, el ingreso residual —que es la ganancia— también crecerá al 5%.

La constancia de la distribución del ingreso también se aplica al periodo de transición. En este periodo el producto por trabajador crece a una tasa superior a la del equilibrio dinámico. Pero la desigualdad se mantendrá constante, independientemente de la tasa de crecimiento. Si la tasa de crecimiento del producto por trabajador en el periodo de transición es de 4% anual, entonces el salario real crecerá también al 4% y la masa salarial y el producto total crecerán al 6%, y por residuo

las ganancias crecerán a esta misma tasa. Aunque las tasas de crecimiento serán mayores, la participación de trabajadores y capitalistas en el ingreso nacional no se modificará.

En el caso de la sociedad omega, la distribución a lo largo de la frontera de crecimiento se mantendrá constante, pues la sociedad omega llegará a ser una sociedad épsilon. En el periodo de transición, el producto por trabajador del sector capitalista crece a una tasa superior a la del sector de subsistencia; luego, la desigualdad aumentará. Una vez que la sociedad omega llega a la frontera de crecimiento, la desigualdad ya no seguirá aumentando y a partir de ese punto se mantendrá constante.

El caso de la sociedad sigma es un poco más complejo. En la etapa de la transición, sigma mostrará un crecimiento económico con aumento en el grado de desigualdad, pues el producto por trabajador en el sector capitalista crece a una tasa mayor que en el sector de subsistencia Y, y este a una tasa mayor que en el sector de subsistencia Z. Una vez que la economía llega a su frontera de crecimiento, quedan solo dos sectores: el capitalista y el de subsistencia Z. El primero crece a tasas más altas que el segundo, pero debido a que el sector capitalista reducirá su tasa de crecimiento en esta etapa, la desigualdad aumentará pero más levemente que en el periodo de transición.

La figura 8.2 presenta las predicciones empíricas del modelo dinámico. En el eje vertical medimos el grado de desigualdad (D) en la distribución del producto total o ingreso nacional —utilizando por ejemplo el coeficiente de Gini— y en el eje vertical el tiempo en años (t). La desigualdad inicial de las tres sociedades está dada por los niveles A (épsilon), B (omega) y C (sigma). En épsilon el grado de desigualdad no se modifica a lo largo del tiempo. En omega, la desigualdad aumenta levemente durante el periodo de transición, hasta llegar a convertirse épsilon (lo que ocurre en el punto J), y a partir de allí se mantiene constante. En sigma, el grado de desigualdad crece continuamente en el periodo de transición, aunque de manera leve, hasta llegar a la frontera de crecimiento, lo cual ocurre en el punto G. A partir de este punto, la desigualdad aumenta aun más lentamente.

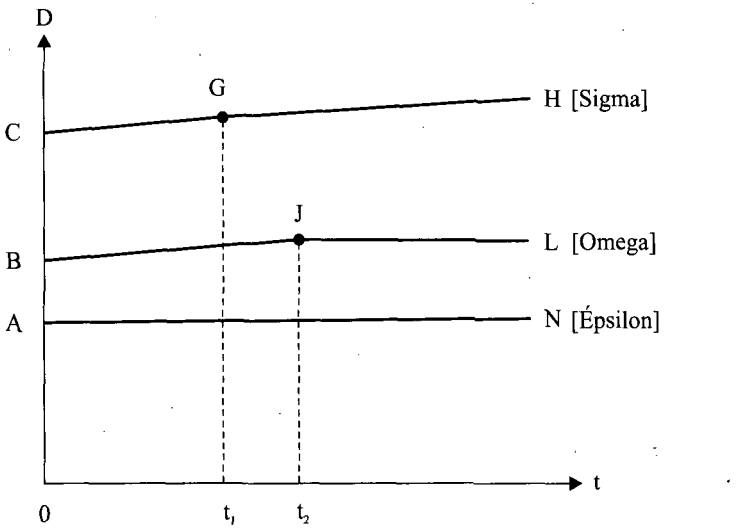

Figura 8.2 Trayectorias de desigualdad de las sociedades capitalistas

La figura 8.2 muestra la trayectoria de largo plazo del coeficiente de desigualdad de cada sociedad capitalista. Las variaciones de corto plazo de esta variable se darán alrededor de la trayectoria de largo plazo. El modelo predice que los cambios en las variables exógenas de corto plazo tendrán el efecto de crear variaciones en el grado de desigualdad alrededor de la tendencia de largo plazo, pero no en la tendencia misma. La trayectoria de largo plazo de la desigualdad en los ingresos en las sociedades capitalistas depende de los factores estructurales de cada tipo de sociedad, como su grado de desigualdad inicial.

Sobre la desigualdad en el capitalismo, este modelo de la teoría unificada predice dos cosas. Primero, que el orden en el grado de desigualdad inicial con el que nacieron las sociedades al capitalismo tenderá a mantenerse a través del tiempo; segundo, y como consecuencia de lo anterior, que no existe tendencia a la convergencia en el grado de desigualdad entre las tres sociedades de manera endógena.

Hay que notar, nuevamente, que las predicciones son sobre niveles de desigualdad entre países, no sobre las variaciones en la desigualdad

entre países. Estas variaciones, como se puede ver en la figura 8.2, dependen de la situación de los países en cuanto a su situación de equilibrio dinámico, si están en el periodo de transición a la frontera de crecimiento o en la misma frontera, y también dependen de los efectos de factores exógenos de corto plazo.

Para la falsación de estas predicciones, el hecho relevante es la sétima regularidad que señala: los países del primer mundo —sociedades tipo épsilon— tienen un grado menor de desigualdad que los del tercer mundo —sociedades tipo sigma— y estas desigualdades persisten en el tiempo. Estas relaciones son precisamente las que aparecen en la figura 8.2. Las predicciones del modelo son consistentes con estas regularidades del capitalismo.

La explicación de la desigualdad de las naciones, según este modelo, es que en cada tipo de sociedad real las desigualdades de hoy tienen su causa en las desigualdades de ayer, en las desigualdades iniciales. Existe un peso de la historia en este proceso. Los países que conforman el primer mundo se iniciaron al capitalismo como países relativamente más igualitarios, comparados con los países del tercer mundo, y estas diferencias se han mantenido en todo este tiempo. El grado de desigualdad, una vez instaurada en una sociedad, tiende a volverse viscosa a través del tiempo. El proceso de crecimiento económico del capitalismo no ha podido eliminar por si solo ese peso de la historia.

Finalmente, este modelo dinámico de la teoría unificada predice que en el proceso de crecimiento económico, los salarios reales crecen. Esta predicción es consistente con la quinta regularidad, del conjunto de regularidades del capitalismo que fueron establecidas en el segundo capítulo.

Conclusiones

Según el modelo dinámico de la teoría unificada presentada aquí, las diferencias en el nivel de desarrollo de los países capitalistas —combinación

de nivel de ingreso con grado de igualdad en la distribución de ingresos— se explican por las diferencias en sus condiciones iniciales. El modelo predice que, en el largo plazo, tanto la trayectoria del nivel de ingreso como la trayectoria del grado de desigualdad en las sociedades capitalistas dependen del grado de desigualdad con que nacieron al capitalismo.

Ahora se puede entender por qué después de tantos años de capitalismo, las diferencias en los niveles de desarrollo se mantienen. Según la teoría unificada, no existe mecanismo en el proceso económico que pueda romper con el legado de la historia de manera endógena. El equilibrio es el que aparece en las figuras 8.1 y 8.2. Se ha mostrado además que la gran mayoría de los países del tercer mundo se parecen a la sociedad abstracta sigma. Los países tipo omega son pocos. En el agregado, por lo tanto, podemos referirnos al tercer mundo como sociedades tipo sigma.

Las regularidades empíricas sobre el desarrollo capitalista, que se establecieron en el segundo capítulo de este libro, no refutan las predicciones del modelo dinámico ni del estático de la teoría unificada. Por lo tanto, no existe razón alguna para rechazar la teoría unificada del capitalismo en esta etapa de nuestra investigación.

La teoría unificada supone al sistema capitalista mundial como compuesto de tres tipos de sociedades abstractas. Pero, tomado en su conjunto, ¿qué tipo de sociedad es el capitalismo mundial? La sociedad capitalista abstracta será el resultado de la agregación de los tres tipos de sociedades abstractas. En cuanto a sus condiciones iniciales, su dotación factorial inicial la hará una sociedad superpoblada y su desigualdad inicial, una sociedad socialmente heterogénea. La agregación de sociedades superpobladas y no superpobladas nos dará una sociedad superpoblada. La agregación de sociedades desiguales en la dotación inicial de activos económicos nos dará una sociedad desigual en esos activos.

Finalmente, la agregación de sociedades socialmente homogéneas y heterogéneas nos dará una sociedad heterogénea. En este caso, las desigualdades en el grado de ciudadanía se dan no solo dentro de las

sociedades sigma, sino entre épsilon y sigma, pues entre ellas existe una historia de relaciones coloniales que genera una jerarquía social. En suma, la sociedad capitalista tomada en su conjunto es una sociedad de tipo sigma. Sabemos cómo funciona una sociedad sigma y podemos, entonces, explicar con esta teoría los hechos básicos del mundo mayor en que vivimos.

Suponer que el mundo capitalista como un todo es una sociedad tipo sigma implica que la desigualdad en esta sociedad mayor debe ser superior que la de los tipos de sociedades que la componen. En efecto, los cálculos hechos sobre la desigualdad mundial —el ingreso total mundial distribuido entre los individuos del planeta— llegan a establecer un coeficiente de Gini del orden de 0.64 para la última década (Milanovic 2006). Esta cifra está por encima incluso de los países más desiguales —Brasil, por ejemplo, está en el orden de 0.58—.

Por otro lado, el fenómeno de las migraciones ilegales hacia el primer mundo está creando allí una masa de población Z. Este hecho sugiere que existe actualmente una tendencia de las sociedades épsilon a devenir en sociedades sigma. Este resultado parece muy extraño a la luz de la teoría unificada presentada aquí, donde las sociedades sigma buscan llegar a ser épsilon; sin embargo, hay que recordar que esta teoría hace abstracción del fenómeno de las migraciones. Habrá que observar si este hecho se vuelve una regularidad empírica del capitalismo y solo entonces habrá la necesidad de explicarla.

Capítulo 9

Naturaleza de las políticas públicas

Existen dos tipos de ciencias en la economía: la ciencia positiva y la ciencia normativa. Hasta aquí la economía ha sido presentada como una ciencia positiva, pues ha buscado explicar *cómo es* el mundo social en que vivimos. Ahora estudiaremos la parte normativa de la ciencia económica, que tratará de responder la pregunta *cómo debería ser* ese mundo social. Esta pregunta pertenece al campo de las políticas públicas, que es el espacio donde los actores sociales proponen y debaten sus objetivos y acciones para modificar la realidad social en cierta dirección.

¿Es realmente necesario recurrir a la ciencia normativa para proponer cómo debería ser el mundo social en que vivimos?, ¿acaso no podemos simplemente proponer lo que nos dicte nuestros deseos? La respuesta es no. La propuesta tiene que tener una lógica interna que la justifique. Por eso se necesita de una ciencia, una ciencia normativa.

La ciencia normativa utiliza el método de las ciencias formales, es decir, deberá construir proposiciones que sean lógicamente correctas, deberá construir teorías normativas. Una teoría normativa se compone de un principio normativo y de una teoría positiva que la haga viable y que se relacionan en un sistema lógico. Luego, la primera tarea consiste en conocer los principios normativos, la ética, que se utiliza en la ciencia económica. La segunda tarea consiste en analizar la consistencia del principio normativo con una teoría positiva que explique el mundo real. Esta segunda condición es menos entendida y tiene que ver con

las interrelaciones lógicas que existen entre teorías normativas y teorías positivas, entre ética y epistemología, en la ciencia económica. Este capítulo desarrolla ambos temas y las políticas públicas para el desarrollo económico constituyen el campo de aplicación.

Implicancias de la teoría unificada para las políticas públicas

El objetivo de una política pública necesita criterios normativos que le sirvan de derrotero. ¿Cuáles son los objetivos que se deben perseguir? Podríamos comenzar utilizando como criterio normativo el campo mismo de la ciencia económica: la producción y la distribución en sociedades humanas. El objetivo puede ser entonces presentado así: se debe buscar el mayor nivel de desarrollo económico de las sociedades. Una sociedad se encuentra en una situación superior de desarrollo económico cuando produce una mayor cantidad de producto por persona y cuando reduce su grado de desigualdad.

El resultado fundamental de la teoría unificada del capitalismo es que el sendero de desarrollo de los países del primer mundo es distinto al de los países del tercer mundo. Las figuras 8.1 y 8.2 así lo muestran. A la luz de esta teoría, el objetivo de política mencionado implica que el tercer mundo debe alcanzar al primer mundo, lo cual a su vez implica que los países del tercer mundo deben moverse hacia el sendero del primer mundo. Este objetivo se diferencia claramente del que proponen las actuales elites políticas y económicas: el tercer mundo debe aumentar su tasa de crecimiento económico, debe crecer más, pues mayor crecimiento reducirá también la desigualdad. Hay que notar que esta propuesta se deriva, implícitamente, de la teoría neoclásica —el paradigma actual en la ciencia económica—, según la cual existe un único sendero en el desarrollo económico capitalista.

Como podemos ver, las propuestas normativas en las políticas públicas no son independientes de las teorías positivas. La relación

existe, sea de manera explícita o implícita. El economista inglés John Maynard Keynes entendió bien esta relación y, por eso sentenció: «Todo hombre público no es sino un prisionero de las ideas de algún economista difunto».

La política pública necesita de instrumentos para conseguir los objetivos deseados. ¿Cómo lograr los objetivos establecidos? Los instrumentos están dados por las variables exógenas de la teoría positiva que se utilice. En la teoría unificada las variables exógenas son las condiciones iniciales de los países (su historia). Aquí se incluyen las dotaciones factoriales y el grado de desigualdad inicial de las sociedades. Ciertamente, las condiciones iniciales con las que partieron los países en el desarrollo capitalista no se pueden cambiar, pero sí se puede romper con el legado de esa historia.

La teoría unificada muestra que un cambio en las dotaciones factoriales en el primer mundo, más cantidad de capital por trabajador, no cambiará su frontera de crecimiento. Como se pudo ver en la figura 8.1, del capítulo anterior, mayor capital por trabajador modificará la trayectoria del ingreso en el periodo de transición de manera exógena (elevará la curva SC), pero el ingreso de destino, la frontera de crecimiento (la curva S^*R^*), seguirá siendo la misma.

En cambio, si se modificara la desigualdad actual en la distribución de activos económicos y políticos en el tercer mundo, el sendero de desarrollo del tercer mundo se movería hacia el del primer mundo (la curva S^*R^* se desplazaría hacia la curva E^*F^*). Si la distribución inicial del tercer mundo llegara a ser igual a la del primer mundo, lo cual implica también modificar la distribución actual de activos a escala mundial, y no solo dentro de cada país del tercer mundo, ambos grupos de países tendrían el mismo sendero de desarrollo. Tendrían la misma frontera de crecimiento y la misma frontera de desigualdad y, por lo tanto, el tercer mundo podría alcanzar al primer mundo en el nivel de desarrollo económico.

Si se mantienen las diferencias en las desigualdades iniciales; si el tercer mundo sigue siendo más desigual que el primero, este no podrá

alcanzarlo. En el tercer mundo, la desigualdad opera como una dotación desfavorable de recursos. Esta es la implicancia de la teoría unificada para las políticas públicas.

En la teoría neoclásica del desarrollo capitalista, en cambio, la variable exógena es el comportamiento de los gobiernos. La acción que propone es la igualación en el medio ambiente económico de los países, donde este medio quiere decir un ambiente muy amigable para la iniciativa privada, donde los gobiernos no interfieran con la libertad económica de los capitalistas. El factor último que explica la falta de desarrollo en el tercer mundo es, según la teoría neoclásica, el comportamiento de sus gobiernos. Si se pudiera modificar la acción de los gobiernos hacia favorecer más la iniciativa privada, las inversiones aumentarían, habría mayor crecimiento del producto y como consecuencia habría mayor igualdad en la distribución del producto (Jones 1998: capítulo 7).

Las reformas liberales que se aplicaron a escala mundial en la década del ochenta y parte de los noventa constituyen un experimento para someter la teoría neoclásica a la falsación. Después de casi dos décadas del inicio de las reformas, la literatura empírica ha llegado a reconocer que las reformas no han generado ni mayor tasa de crecimiento ni una reducción de la desigualdad, como muestras las regularidades 6 y 7 del segundo capítulo. Los datos de la realidad refutan la teoría neoclásica del desarrollo.

Como mostramos en el capítulo 8, la teoría unificada del desarrollo capitalista no es refutada por los datos de la realidad; se mostró allí que sus predicciones son consistentes con esas regularidades. A diferencia de la teoría neoclásica, la teoría unificada supone que el comportamiento de los gobiernos es endógeno. El comportamiento de los gobiernos del tercer mundo no son causa sino consecuencia de la falta de desarrollo. No podrían por lo tanto ser los agentes del desarrollo económico, los transformadores de la sociedad. Es por eso que ninguna de las políticas de desarrollo que se han aplicado en el tercer mundo ha dado resultado, pues se han basado en una teoría equivocada. El principio «no hay cosa más práctica que una buena teoría» se hace presente nuevamente.

Según la teoría unificada, los países del tercer mundo están atrapados en un equilibrio de bajo nivel. Los gobiernos son incompetentes, la corrupción es extendida, la ilegalidad es enorme, la inestabilidad política es grande, la inversión no es sostenida, la educación es baja, el ingreso medio es bajo y la desigualdad es alta. Todas estas variables son, según la teoría unificada, endógenas. La variable exógena está en su historia, en el legado de esa historia, que es su desigualdad inicial. El instrumento para salir de la trampa, del equilibrio de bajo nivel, está en reducir la desigualdad inicial, pero no solo la que existe dentro de los países del tercer mundo, sino la que se da entre los países del primer mundo y del tercer mundo, la que se da en el capitalismo mundial.

Según la teoría unificada, la igualación en el grado de desigualdad entre países significaría colocar a los países en condiciones iguales para competir en la economía mundial. En último análisis, dice esta teoría, las sociedades compiten con su grado de desigualdad, con su calidad de sociedad.

Esta variable exógena no ha sido tomada en cuenta en ninguna de las propuestas conocidas sobre políticas de desarrollo económico. Se han propuesto políticas de igualación entre países en otras variables, tales como tarifas al comercio, tipos de cambio real, tasas de interés real, igualdad en los incentivos tributarios para los inversionistas, pero nunca la igualación en los grados de desigualdad.

Pero, ¿en qué consistiría cambiar la desigualdad inicial? La teoría unificada considera dos tipos de activos: activos económicos —tierra, capital físico, capital humano— y activos políticos —ciudadanía—. Se trata de reducir la desigualdad en la distribución de todos estos activos. Una reforma agraria que redistribuye tierras es parte de la política, pero no es todo; una redistribución solo del capital físico tiene la misma característica. Los efectos limitados de los programas de reforma agraria, que redistribuyeron tierras, pero ningún otro activo más, sugieren que el capital humano y la ciudadanía son los fundamentales.

La teoría unificada muestra que la mayor igualdad en el capital humano y en la ciudadanía van muy juntos, pues el proceso educativo

depende del grado de democracia del país (capítulo 7). Y en el largo plazo ambos se refuerzan. El mayor grado de ciudadanía llevaría a las masas a demandar una educación de mayor calidad y así a una mayor igualdad en el capital humano. La mayor igualdad en el capital humano llevaría a las masas a demandar igualdad de ciudadanía. Luego, la igualdad en ciudadanía llevaría a una mayor igualdad en el capital humano y así sucesivamente.

Como se puede apreciar, cambios en la desigualdad inicial implican acciones complejas. Van más allá de las políticas convencionales de combate a la pobreza, donde los gobiernos buscan transferir solo dinero a los pobres de varias maneras. Esta política supone que la única diferencia entre ricos y pobres es la cantidad de dinero o de ingresos; y que en el resto de los activos son iguales. Se aplican políticas que suponen sociedades socialmente homogéneas a sociedades que no lo son.[8]

En las sociedades tipo sigma, que como hemos mostrado aquí incluyen tanto el tercer mundo como el capitalismo mundial, la política para cambiar la desigualdad de ingresos consiste en cambiar la distribución de los activos económicos y políticos. Cambios en la distribución de activos implica un rompimiento con la historia. Son medidas que implican un episodio refundacional. Los países del tercer mundo no han roto con su pasado, pues las desigualdades iniciales se reproducen en el tiempo. Se puede decir que estos países no tienen historia, pues la historia está todavía presente. Un conocido escritor mexicano dijo «México no tiene historia hasta ahora». Pero igual se puede decir del capitalismo mundial, que es también una sociedad tipo sigma. El capitalismo mundial no ha roto con su historia. La desigualdad inicial entre países se reproduce actualmente. El peso de la historia cuenta hasta hoy en el proceso de desarrollo económico del capitalismo.

[8] El famoso diálogo entre los escritores americanos Hemingway y Fitzgerald, donde concluyen que «la única diferencia entre los ricos y los pobres es que los ricos tienen más dinero», tenía como contexto ciertamente una sociedad épsilon.

La reducción en el grado de desigualdad de la ciudadanía, en particular, es un cambio cualitativo que haría que un país tipo sigma tuviera historia, pues esa desigualdad ya pertenecería al pasado, a la historia. Por eso estos cambios implicarían un episodio refundacional en la historia de los países del tercer mundo. De igual manera, se puede decir que la reducción de la desigualdad en ciudadanía mundial entre los países del primer mundo y del tercer mundo significaría un episodio refundacional en el capitalismo mundial.

La pregunta ahora es, ¿quiénes serían los actores sociales que tendrían el poder y el incentivo para llevar a cabo una política refundacional? En la teoría unificada hemos mostrado que el equilibrio económico al que llegan las sociedades implica que ningún actor social tiene ni el poder ni el incentivo para cambiar la situación; por lo tanto, la situación de equilibrio —sea estática o dinámica— se repetirá periodo tras periodo. Según la teoría unificada, los actores sociales que hemos estudiado —capitalistas, gobiernos, trabajadores— no podrían ser los agentes de cambio. Las élites económicas y políticas tienen el poder pero no tienen los incentivos; por el contrario, los trabajadores tienen el incentivo pero no tienen el poder. Estos comportamientos y sus limitaciones ya están incorporados en el equilibrio económico.

La conclusión que emerge de la teoría unificada es inevitable. El agente del cambio social, el empresario social del desarrollo, solo puede surgir de manera exógena. Pero queda todavía la pregunta, ¿qué conclusiones se pueden extraer de las teorías normativas que existen en la ciencia económica?

Principios normativos

El objetivo de una teoría económica normativa es establecer criterios o principios éticos que deben guiar la evaluación sobre cuán bien funciona una sociedad, así como las acciones que se deberían tomar para mejorarla. Hasta aquí hemos utilizado un criterio normativo que se

deriva de la teoría unificada, las implicancias normativas de la teoría positiva. Pero existe en la literatura principios de bienestar social que bien vale la pena resumir aquí. Los más conocidos son los siguientes.

Principio de Pareto

Wilfredo Pareto fue un economista italiano cuyas publicaciones aparecieron a inicios de 1900. Según el principio de Pareto, una sociedad funciona bien cuando no es posible mejorar el bienestar de alguien sin desmejorar el bienestar de otro. ¿Cuál es la lógica de este principio? Esta teoría normativa proviene de una teoría positiva: la teoría neoclásica.

La teoría neoclásica incluye una teoría de equilibrio microeconómico sobre el comportamiento de los consumidores, que se llama la teoría de la utilidad ordinal. La teoría supone que los individuos tienen una función de utilidad, la cual buscan maximizar. La utilidad del individuo depende de la cantidad de bienes que consume. Los individuos pueden ordenar las canastas de bienes que enfrentan de acuerdo a su función de utilidad y así determinar las preferencias por una canasta en lugar de otra. Los individuos pueden ordenar solamente, pero no pueden cuantificar la magnitud de su utilidad. No existe un hedonímetro incorporado en los consumidores.

Pero, entonces, no se puede comparar las utilidades entre individuos; más aún, no se puede saber el resultado de una agregación, donde un individuo pierda y otro gane. Así, si se transfiere ingresos de una persona rica a otra pobre, no se puede decir si en el agregado la suma de las utilidades ha aumentado, disminuido o se mantiene constante. Para el que recibe ha aumentado y para el que pierde ha disminuido, pero no es posible sumar las utilidades.

La consecuencia es que solo puede haber una situación superior en la sociedad si alguien puede ser mejorado sin desmejorar a nadie. En este caso habrá una clara ganancia en la utilidad agregada, aunque la utilidad no sea mensurable. ¿Cómo puede ocurrir tal situación? Tiene que haber ineficiencia en la sociedad, algo así como «dinero tirado en

la calle por recoger». Tiene que haber oportunidades de producir más bienes que no se están aprovechando, así la mayor cantidad de bienes se puede utilizar para aumentar el ingreso de unos y compensar a los que pudieran perder.

De nuevo, ¿cómo puede ocurrir tal situación? La organización de la producción es ineficiente; la sociedad está desperdiciando sus recursos, pues podría producir más cantidad de bienes con la misma cantidad de recursos que posee. La economía no está produciendo en su frontera de posibilidades de producción.

Por tercera vez, ¿cómo puede ocurrir tal cosa? En este punto, el principio paretiano hace un supuesto sobre el funcionamiento del mundo real: supone que la sociedad capitalista funciona como si fuese una economía neoclásica. La teoría normativa se apoya en una teoría positiva. Según la teoría neoclásica, si los mercados son de competencia perfecta —donde nadie tiene poder de imponer precios— y ningún actor social crea externalidades a los demás —daños o beneficios indirectos por los que no paga o cobra—, la solución de equilibrio general es un óptimo paretiano. La implicancia es que si existen monopolios, externalidades o bienes públicos, la solución de equilibrio general no es un óptimo paretiano, es decir, es ineficiente.

Las acciones que se derivan del principio paretiano es buscar eliminar los monopolios y las externalidades, pues así se podría producir más cantidad de bienes con la misma cantidad de recursos. Las sociedades deben abrir sus economías al mercado internacional para eliminar los monopolios y deben gravar con impuestos a los que producen externalidades negativas. Las ganancias en eficiencia de esas medidas servirían para mejorar a unos sin desmejorar a otros, pues el mayor producto se puede utilizar para compensar a los perdedores, si los hubiera.

El principio paretiano es un principio utilitarista. No es un principio de justicia distributiva. Es un principio que ignora las desigualdades de ingresos entre los individuos. Si se transfiriera ingresos de los más ricos a los más pobres de la sociedad a través del presupuesto público, no habría una situación de pareto superior en la sociedad.

Pues las ganancias y pérdidas en utilidad no se pueden sumar. Para mejorar la situación de los pobres se tendría que mejorar la eficiencia económica de la sociedad.

Principio de Maslow

Abraham Maslow, psicólogo americano, afirmó que, una sociedad funciona bien cuando no existen pobres. Y si existen, la sociedad estará en una situación superior si se mejora a los pobres, aunque para ello se tenga que desmejorar a los ricos.

¿Cuál es la justificación de este principio? Esta teoría normativa se apoya en una teoría positiva: el comportamiento humano se guía por la motivación de satisfacer las necesidades humanas, las cuales tienen un orden jerárquico, de las más importantes a las menos importantes. En términos gruesos, según Maslow, existen las necesidades primarias —que se refieren a la supervivencia física del individuo—, luego vienen las secundarias —que se refieren a las que son necesarias para convivir en sociedad— y finalmente, se encuentran las necesidades de desarrollo humano —deseos por arte, ciencia, política—. Existe una pirámide de necesidades, con las necesidades primarias en la base y las de desarrollo humano en la cúspide, y las necesidades secundarias en el medio. Los individuos buscan llegar hasta la cúspide de la pirámide y si no lo logran es debido a las restricciones que enfrentan. El individuo es una máquina de progreso. Su pobreza es una situación socialmente impuesta, no es su elección.

La teoría positiva de Maslow supone que los individuos tienen funciones de utilidad pero distintos a los que supuso Pareto. Las utilidades de los individuos reflejan la estructura y jerarquía de las necesidades humanas, pues existen necesidades primarias, necesidades secundarias y así sucesivamente. Luego, los individuos ordenan las canastas de consumo de acuerdo a las necesidades que satisfacen. Sus preferencias son lexicográficas, es decir, «primero es lo primero». Si los individuos tuvieran que elegir entre alimentos y turismo, elegirían alimentos —que

satisface una necesidad primaria—; o elegirán turismo —una necesidad de orden superior—, solo si ya han satisfecho la necesidad fisiológica de alimentos.

El ingreso real de los individuos determinará qué necesidades son satisfechas y qué otras quedan sin ser satisfechas. Los pobres son aquellos que con su ingreso no pueden satisfacer ni siquiera sus necesidades primarias. Luego, si se trasfiriera ingresos de los ricos a los pobres, la sociedad estaría en una situación superior.

De esta teoría positiva de Maslow se ha derivado una teoría normativa, a la que hemos denominado el principio de Maslow. De acuerdo a este principio, toda transferencia de ingresos de ricos a pobres eleva la calidad de la sociedad. Pero una transferencia de ingresos entre ricos no mejorará la sociedad; tampoco lo haría una transferencia de ingresos entre pobres. Y por supuesto una transferencia de ingreso de pobres a ricos desmejora la sociedad.

El principio de Maslow es un principio cuasi-utilitarista, pues incluye la utilidad que derivan los individuos de los bienes, pero también hace una prelación entre los bienes, según la necesidad que satisfagan. Es, además, un principio donde se puede justificar la mayor desigualdad solo si implica mejorar el ingreso de los más pobres. Es un principio de justicia distributiva. La situación óptima de Maslow implica la igualdad en los ingresos en la sociedad, donde ya no se necesita hacer transferencias de ingreso.

Principio de Rawls

John Rawls, filósofo estadounidense, estableció un principio que lleva su nombre. Allí afirma que una sociedad funciona bien si los individuos pueden contar con los bienes primarios que se necesitan en la sociedad, si la desigualdad es mínima. ¿Cuál es la lógica de este principio? Esta teoría normativa se apoya en una teoría positiva: la teoría del contrato social.

La teoría del contrato social supone que en la sociedad los individuos no conocen su situación futura, tienen un velo de ignorancia sobre su futuro y están, por lo tanto, interesados en establecer unas reglas distributivas desde el inicio. Estas reglas constituyen el contrato social. Bajo estas circunstancias, las reglas que establece dicho contrato serán objetivas y equitativas.

El contrato social incluye dos reglas ordenadas lexicográficamente: (a) todos deben tener un ingreso mínimo que les permita consumir los bienes primarios de la sociedad; (b) la desigualdad no debe pasar ciertos límites, de modo que cuando algún individuo obtenga ingresos extraordinarios estos sean compartidos por el resto. Estas reglas constituyen un contrato social que será considerado justo por todos.

La teoría del contrato social hace supuestos adicionales sobre el mundo real. Supone que en la sociedad existen los bienes primarios y que estos son de carácter social; asimismo, supone que los individuos tienen una tolerancia limitada a la desigualdad. El primer supuesto es similar al de la teoría positiva de Maslow y el segundo es similar al de la teoría positiva de la tolerancia limitada a la desigualdad que desarrollamos en el quinto capítulo.

Cuando en la sociedad existen pobres, individuos con ingresos por debajo del mínimo, el criterio de bienestar indica que el objetivo es aumentar el ingreso de los pobres, como en el caso del principio de Maslow. También en este caso se puede justificar un aumento en la desigualdad si con ello se eleva el ingreso de los pobres. Cuando todos ya estén por encima del ingreso mínimo, el criterio de bienestar indica que el objetivo ahora es reducir la desigualdad, pero no totalmente sino hasta que sea socialmente aceptada. El principio de Rawls es también de justicia distributiva.

Principio de Sen

Amartya Sen es un economista y filósofo indio. Según el principio propuso, una sociedad funciona bien cuando existe igualdad de capacidades

entre los individuos. ¿Cuál es la lógica de este principio? Los bienes no importan mucho por sí mismos, sino por lo que pueden hacer por la gente. Los bienes permiten a la gente funcionar (movilizarse, trabajar, estudiar y aprender) y lo que importa es que la gente funcione lo mejor que sea posible. Son estas capacidades las que deben ser maximizadas en la sociedad y las igualdades en capacidades las que se deben buscar.

El principio de Sen también se apoya en una teoría positiva: la teoría de las capacidades. Hace supuestos implícitos sobre el mundo real. Supone una jerarquía de necesidades, entre las que construyen capacidades y las que no lo hacen; además, entre las primeras debe existir una jerarquía de bienes, pues los bienes hacen que los individuos puedan funcionar de varias maneras y algunas de esas funciones pueden ser más importantes que otras; es decir, ciertos bienes están asociados a funcionamientos específicos. Los bienes y los funcionamientos, así como las capacidades para llevar a cabo los diversos funcionamientos, están jerarquizados. El principio de Sen contiene así elementos de la teoría positiva de Maslow.

También este es un principio de justicia distributiva y de igualdad de oportunidades. Pero esta igualdad de oportunidades va más allá de las igualdades formales, en donde lo que se busca igualar no son necesariamente los ingresos, sino las capacidades de los individuos. Individuos que nacieron en ambientes que los llevaron a tener capacidades limitadas —como los discapacitados— deben ser compensados y recibir más ingresos que el promedio para llegar a las mismas capacidades de los demás. Supone una sociedad sigma buscar la igualdad de capacidades, en lugar de la igualdad de ingresos, es la esencia del principio de Sen.

Integración entre teorías normativas y positivas

Una teoría normativa es, tal como la definimos antes, un sistema lógico que relaciona un principio normativo con una teoría positiva que le sirve se sustento. Si una teoría normativa va a servir para la política

pública tiene que apoyarse en una *buena* teoría positiva. Existe, por lo tanto, una causalidad implícita en una teoría normativa. La causalidad solo puede provenir de la teoría positiva, donde la variable endógena de la teoría positiva es el objetivo buscado y la variable exógena es el instrumento para llegar al objetivo.

Sea X el instrumento y E el objetivo. La interrelación entre teoría normativa y teoría positiva se puede escribir así:

> Si el agente social A mueve la variable X en cierta dirección,
> la variable normativa E alcanzará un valor superior

Esta proposición tendría que cumplirse en la realidad si la teoría normativa va a tener alguna utilidad práctica en el diseño de las políticas públicas. Esta proposición es, en efecto, falsable. La única parte de la proposición que no es falsable es el principio normativo, el principio ético, que establece que mayores valores de E son deseables; el resto es falsable (que E depende de X y que el actor social A tiene el poder y el incentivo para modificar X). La teoría normativa puede, en consecuencia, tener el problema de estar apoyada en una teoría positiva que es empíricamente falsa. En este caso, la proposición normativa, en su conjunto, será también empíricamente falsa.

La proposición normativa puede fallar debido a las razones siguientes:

(a) La lógica del principio normativo se basa en una teoría positiva y esta puede ser empíricamente falsa, es decir, cambios en X puede no generar el valor de E en la dirección deseada.

(b) El actor social que debe mover la variable X en una dirección determinada puede no tener el poder y los incentivos para hacerlo.

Sobre el problema «a», hemos visto anteriormente que el principio de Pareto supone una teoría microeconómica de la utilidad ordinal y escalar, mientras que el principio de Maslow supone una teoría

lexicográfica. La evidencia empírica apoya la segunda, pues los estudios empíricos muestran que existe una jerarquía de bienes en las canastas de consumo de ricos y pobres. En ese sentido, el criterio de Pareto quedaría invalidado para la aplicación práctica.

Los dos supuestos que subyacen al principio de Rawls tienen consistencia empírica. Para la primera, ya indicamos la existencia empírica de la jerarquía de bienes en las canastas de consumo de los individuos; y para la segunda, los estudios empíricos muestran que la excesiva desigualdad no es tolerada socialmente y causa desorden social, como se mostró en el capítulo 5. Finalmente, el principio de Sen tiene consistencia con la existencia empírica de una jerarquía en los bienes.

La teoría positiva que sirva de apoyo a la teoría normativa no solo debe tener una buena teoría microeconómica sino también una buena teoría macroeconómica. El principio de Pareto supone una sociedad neoclásica. Hemos mostrado en este libro que los datos de la realidad sobre producción y distribución refutan la teoría del equilibrio general neoclásico. En los otros casos, el principio normativo no supone una teoría económica de equilibrio general particular y entonces se les podría integrar a la teoría unificada del capitalismo. Los principios de Maslow, Rawls y Sen son principios de justicia distributiva, no son principios de bienestar.

Sobre el problema «b», en todos los principios descritos, el actor social implícito es el gobierno. Pero los gobiernos pueden no buscar ni la eficiencia económica, ni la igualdad en la distribución del ingreso, ni un piso de ingresos en la sociedad, ni la igualdad de las capacidades. En efecto, la teoría de los gobiernos propuesta en este libro (capítulo 5) supone que los gobiernos actúan guiados por sus propios intereses, y las predicciones de la teoría parecen ser consistentes con los hechos observados. Esta teoría predice que los gobiernos no tienen suficiente poder ni los incentivos para llevar a cabo las acciones que establecen los principios normativos presentados aquí.

Así como mostramos en el capítulo sobre epistemología (capítulo 1) que una teoría positiva requiere de una teoría normativa —epistemológica—, ahora tenemos el caso de que una teoría normativa

—ética— requiere de una teoría positiva sobre la realidad que desea influir. Existe, por lo tanto, una conexión entre las dos ramas de la filosofía, la epistemología y la ética, con la ciencia económica.

¿Cuál es la teoría normativa que hemos utilizado implícitamente en la política de desarrollo económico propuesta anteriormente?

El principio de la igualdad de oportunidades es común a las tres teorías normativas, pues en Maslow se busca igualar las oportunidades por medio de la igualación en la satisfacción de las necesidades humanas; en Rawls, por medio de los bienes sociales primarios y una desigualdad limitada; y, en Sen, por medio de la igualación de las capacidades. Igualdad de oportunidades en este sentido es un concepto más radical que el puramente formal que se usa en el lenguaje corriente. Las constituciones de los países del tercer mundo, que es donde se expresa el contrato social, contienen disposiciones a favor de la igualdad de oportunidades, pero son usualmente disposiciones formales solamente, sin efectos prácticos. Así, existe una brecha entre el país formal y el país real.

Si tomamos para las políticas públicas el principio normativo de la igualdad de oportunidades, la teoría unificada nos dice que ese objetivo implica modificar las condiciones iniciales de los individuos y las sociedades. Es una sociedad tipo sigma, la igualdad de oportunidades tiene que ver con la distribución de los stocks y no de los flujos. En consecuencia, el objetivo de la política pública que se adoptó en la primera parte de este capítulo —el nivel de desarrollo de las sociedades— y el instrumento que se propuso —la redistribución de los activos económicos y políticos— es consistente con este principio.

Nuevas implicancias para las políticas públicas

Hasta ahora hemos ignorado el factor del medio ambiente en el proceso económico. El modelo dinámico de la teoría unificada predice que el crecimiento económico de los países del primer mundo y del tercer

mundo, aunque siguen senderos diferentes, puede repetirse periodo tras periodo; es decir, predice que el crecimiento económico puede seguir indefinidamente.

Esta proposición es contraria a las leyes que establecen la física y la biología; en particular, la Ley de la Entropía predice que existen límites al crecimiento económico. Como se muestra en el apéndice, el proceso económico ignora los factores ambientales. Primero, el proceso económico produce no solo bienes, sino también males, como la cantidad de desperdicio que sale de dicho proceso. A mayor producción de bienes, mayor producción de desperdicio y mayor grado de contaminación al medio ambiente. Segundo, en el proceso económico ingresan no solo cantidades de capital y trabajadores, sino también recursos naturales. Los primeros tipos de stocks pueden reproducirse mediante la depreciación y los ingresos de subsistencia; los segundos, en cambio, no pueden reproducirse, son agotables. Si la *misma cantidad* de bienes se produce periodo tras periodo, si el flujo producido de bienes es el mismo, el stock de recursos naturales disminuirá y la capacidad productiva del sistema disminuirá *constantemente*; por lo tanto, si se aumenta el flujo de bienes producidos (si se busca el crecimiento económico), el deterioro de la capacidad productiva del sistema se amplificará.

Los humanos, vistos como especie biológica, enfrentan una frontera de posibilidades de crecimiento intergeneracional. Cuanto más cantidades de bienes produce la presente generación de humanos, menos bienes estarán disponibles para las futuras generaciones. La producción de ahora reduce la capacidad productiva de la madre naturaleza para mañana, de manera irrevocable.

Ciertamente, un país individual puede crecer indefinidamente. Pero, como ocurre a menudo con el análisis económico, existe el principio de la falacia de composición: si todos los países quisieran crecer indefinidamente, no podrían hacerlo. El biólogo Edward Wilson ha reportado un cálculo fundamental. Si todos los países del tercer mundo quisieran alcanzar el nivel de consumo actual de los Estados Unidos, con la tecnología actual, los recursos de la Tierra no

alcanzarían y se necesitarían dos planetas Tierra adicionales. Si por la vía del crecimiento económico el tercer mundo no podrá nunca alcanzar el nivel de vida de los países del primer mundo, este objetivo se tendría que abandonar por inviable en términos físicos.

El modelo dinámico de la teoría unificada ha hecho una abstracción del factor ambiental. Ciertamente, se puede construir un modelo nuevo que incluya esta restricción en el proceso económico. Este modelo tendría como nueva variable exógena las restricciones que establece la Ley de la Entropía y como nueva variable endógena la distribución de los bienes entre generaciones. En realidad, tales modelos ya han sido construidos (Georgescu-Roegen 1971), pero no son parte de la economía estándar.

La introducción del medio ambiente en la teoría unificada tiene consecuencias enormes para las políticas públicas. El objetivo de buscar el mayor nivel de ingreso medio de nuestra generación ya no podría constituir el objetivo único. Así, el objetivo no podría ser el de maximizar la tasa de crecimiento del producto, como hoy día proponen las elites políticas y económicas. En los objetivos se tiene que incluir las necesidades de las siguientes generaciones. Será necesario establecer un nuevo principio normativo que tome en cuenta las necesidades de las próximas generaciones. Se necesita un principio ético para resolver el dilema intergeneracional: más bienes para nuestra generación implica menos para las siguientes. En consecuencia, el manejo del medio ambiente tendría que ser parte del objetivo del desarrollo económico.

En cuanto a los instrumentos, también las consecuencias son importantes. El objetivo de buscar la igualdad en el nivel de ingresos entre el primer mundo y el tercer mundo por la vía del crecimiento económico, como propone la economía estándar, no sería físicamente viable. El instrumento sería la redistribución de los flujos de ingresos del primer mundo al tercer mundo. Dado que el capitalismo mundial es una sociedad tipo sigma, esa redistribución de flujos implica una previa redistribución de activos también a escala mundial. La igualación de la ciudadanía a escala mundial es entonces uno de los

instrumentos fundamentales para lograr tal objetivo. La globalización actual tanto del comercio, del capital, del trabajo, así como el de las comunicaciones, constituye tal vez una rendija para lograrlo. Pero, nuevamente, el problema del actor social que pueda llevar a cabo esas medidas solo puede aparecer exógenamente.

Debemos considerar ahora otro factor que la teoría unificada ha ignorado. Los modelos de la teoría unificada estudiados hasta aquí suponen que existe movilidad del capital entre sociedades, pero no existe movilidad de trabajadores. La inmigración internacional de trabajadores del tercer mundo al primer mundo, que ocurre actualmente con mayor fuerza, ha sido ignorada por la teoría unificada. El supuesto implícito ha sido que la inmigración no es un factor esencial en el funcionamiento del capitalismo mundial; es decir, la inmigración no tiene el efecto de desplazar el sendero de desarrollo del tercer mundo hacia el del primer mundo.

Pero es probable que las migraciones internacionales tengan una mayor importancia en el futuro. La inmigración actual es mayormente de trabajadores de poco capital humano y hacen trabajos que los trabajadores nativos del primer mundo ya no quieren realizar; es, además, una inmigración ilegal. Esta inmigración está generando un grupo de trabajadores que son ciudadanos de segunda categoría. En el primer mundo está creciendo el grupo de trabajadores Z. Así, las sociedades épsilon parecen tener una tendencia a convertirse en sociedades sigma.

De acuerdo a la teoría unificada, un mundo muy desigual es un mundo que funciona con desorden social, y el desorden social y la violencia tienen su origen en la excesiva desigualdad. Luego, el grado de desorden social y de violencia que se observa en el mundo se origina en la pronunciada desigualdad mundial. Este es el costo social de la desigualdad: la calidad de vida de todos se degrada. Las hipótesis alternativas señalan que la violencia mundial se debe a diferencias en la cultura, especialmente en la religión. Pero ambas son complementarias y puede entonces deberse a ambos factores.

Según la teoría unificada, para reducir el nivel de desorden social del mundo en que vivimos se debe reducir el grado de desigualdad que existe a escala nacional y mundial. El objetivo de reducir el desorden social también tiene un instrumento para lograrlo: la reducción en la desigualdad inicial que existe tanto dentro de los países del primer mundo como entre el grupo de países del primer y tercer mundo.

También según la teoría unificada, la desigualdad en los activos políticos juega un papel importante en el proceso económico. Si en cada país del tercer mundo se tuviera la misma ciudadanía, la sociedad funcionaría de otra manera. Los derechos y obligaciones serían los mismos. La democracia sería más fuerte, pues los gobiernos tendrían que responder a las demandas de bienestar de las masas y tendrían que darles cuenta de sus políticas. Las masas formarían parte de los grupos de interés que imponen restricciones en el comportamiento de los gobiernos. Las firmas tendrían que responder a las demandas de bienestar no solo de sus trabajadores sino a las de la sociedad, como por ejemplo el cuidado del medio ambiente y de los estándares laborales. Un sistema económico compuesto de trabajadores que son ciudadanos de primera categoría sería no solo más democrático, sino más productivo e igualitario. La sociedad en su conjunto sería de mejor calidad para todos.

Si todos tuvieran la misma ciudadanía a escala global, el mundo sería un mejor lugar para vivir. Todos ganarían. Sería una situación superior en términos de los más importantes criterios de justicia distributiva que ha desarrollado la economía normativa: superior en términos de Maslow, Rawls y Sen.

Por contraste a la teoría unificada, las implicancias de la teoría neoclásica para las políticas públicas son totalmente distintas. El principio que utilizan es el criterio de Pareto. Como hemos visto, esta teoría se apoya en una teoría positiva errada, que no explica el funcionamiento del capitalismo. La teoría neoclásica hace abstracción de la importancia de la desigualdad inicial; en particular, abstrae las diferencias en activos políticos. Supone que toda sociedad capitalista, sea nacional o mundial, es una sociedad del tipo épsilon u omega. Solo hay ricos y pobres en

el mundo. La teoría neoclásica abstrae el activo político en el proceso económico; las sociedades capitalistas tipo sigma no existen. Pero es una abstracción equivocada, pues hace abstracción de un factor que es esencial para entender el funcionamiento de la sociedad capitalista, nacional o mundial. Por eso falla la teoría neoclásica.

El principio de Gödel en la ciencia económica

Según la teoría unificada del capitalismo, el factor último que explica el atraso relativo del tercer mundo es la desigualdad inicial en los activos económicos y políticos, la historia. Entre ellos, como hemos visto anteriormente, la ciudadanía parece tener un papel significativo. ¿Cómo crear igualdad de ciudadanía dentro de los países del tercer mundo y a nivel global para mejorar el mundo en que vivimos?, ¿quiénes serían los actores sociales con poder e incentivos para hacerlo?

La respuesta que emerge de la teoría unificada es que tales cambios no podrían ocurrir endógenamente. En el equilibrio general, estático o dinámico, la solución de producción y distribución se repite periodo tras periodo, mientras los valores de las variables exógenas se mantengan fijos. La situación de equilibrio significa que ninguno de los actores sociales tiene el poder y el incentivo para cambiar la situación. El cambio, entonces, tendría que venir de afuera del proceso económico, de manera exógena.

Esta conclusión puede parecerle muy extraña, amigo lector. ¿Cómo es que los problemas económicos no pueden resolverse dentro de la economía? Este problema no es nada nuevo, se le conoce en matemáticas como el *problema de Gödel.* Kurt Gödel fue el matemático estadounidense de origen austriaco que desarrolló dos teoremas importantes en 1931. La implicancia de estos teoremas es que existen sistemas que no pueden explicarse así mismos. Fuera de las matemáticas, se puede citar varios ejemplos de este problema: ¿quién es el juez de los jueces?, ¿quién es el supervisor de los supervisores, el evaluador de los evaluadores?,

¿cuál es la teoría de las teorías?, ¿cuál es el principio de los principios?, ¿cuáles son las palabras que definen las palabras de un diccionario?

Otro ejemplo se da en la ciencia económica. Según la teoría unificada *el proceso económico no puede explicarse así mismo.* Se necesitan variables exógenas que lo expliquen. La implicancia lógica es que el *proceso económico no puede modificarse así mismo.* Si un problema económico pudiese resolverse dentro del proceso económico, no existirían problemas económicos por resolver. La sociedad tendría la capacidad de autorregularse, de resolver sus problemas de manera endógena. La observación empírica nos muestra que tenemos problemas sociales no resueltos y de larga data, nos indica que en la economía estamos enfrentados al problema de Gödel.

La solución del problema económico está fuera del proceso económico. Los actores sociales que podrían modificar las variables exógenas del sistema capitalista aparecerán de manera exógena. Pero tampoco la solución del problema económico estará en las otras ciencias sociales. Debido a que las ciencias sociales se interrelacionan, al final, tomadas como un sistema, no podrán comprenderse a sí mismas. Por otro lado, para un país dado, la variable exógena puede provenir de otro país, pero para el conjunto de países, para el sistema capitalista mundial, la variable exógena no puede provenir de ella misma; de igual manera, para un grupo social dado, la variable exógena puede prevenir de otro grupo social, pero para el conjunto de grupos, para la sociedad, la variable exógena no puede provenir del proceso económico de ella misma. La falacia de composición en acción aparece nuevamente.

Este libro no va a concluir, como hacen otros libros de economía, con un listado de acciones que se deben hacer —¿por quién?— para cambiar el mundo social en que vivimos. Tampoco va a pasarles ese listado a los gobiernos, pues su comportamiento es determinado endógenamente. Las acciones que podrían escapar al problema de Gödel se indicarán en el capítulo siguiente, en las conclusiones.

Capítulo 10

Conclusiones

Este capítulo presenta los principales resultados del libro. Debemos ahora retomar las preguntas formuladas al inicio del libro y darles respuesta.

¿Qué debemos saber primero para entender la realidad social?

La ciencia económica se ocupa del problema de la producción de bienes y su distribución entre los grupos sociales en las sociedades humanas. Para que este problema sea objeto de estudio científico, como se ha mostrado en el primer capítulo, la actividad humana de la producción y distribución tiene que transformarse en un proceso, llamado el proceso económico. Además, para explicar los determinantes de la producción y distribución, la ciencia económica utiliza métodos científicos de investigación. La lógica de la investigación científica que se ha utilizado en este libro es la epistemología popperiana.

La ciencia económica construye teorías positivas para explicar la realidad. Sin teoría no hay explicación posible. Una teoría económica es un artificio lógico que construye el investigador con el fin de explicar los determinantes de la producción y distribución en sociedades humanas particulares. El artificio lógico consiste en la construcción de

una sociedad abstracta y simple que constituya una buena aproximación a las sociedades reales y complejas, que son materia de estudio. Para llevar a la prueba empírica esta aproximación se construyen modelos de la teoría, donde los modelos son formas más específicas que adopta la sociedad abstracta, formas que la hacen empíricamente refutables. La bondad de la aproximación se hace, entonces, contrastando las predicciones empíricas de los modelos de la teoría contra los datos de la realidad.

Para responder la pregunta del subtítulo, para entender la realidad social se necesita conocer las reglas del conocimiento científico. Se necesita conocer una epistemología de manera explícita y aplicarla.

La sociedad capitalista ha sido el objeto de estudio en este libro. Y, la epistemología popperiana ha sido utilizada para tal efecto. Se han presentado las principales teorías que intentan explicar la producción y distribución bajo el capitalismo. Luego, las predicciones empíricas de las teorías han sido sometidas a la prueba de la falsación. Los datos de la realidad que se han utilizado para tal efecto han sido establecidos como un conjunto de regularidades empíricas de los países capitalistas, regularidades que también se refieren a la producción y distribución.

¿En qué mundo social vivimos?

El mundo capitalista ha sido dividido en dos grupos de países siguiendo el criterio del nivel de ingreso: el primer y el tercer mundo, donde el primer mundo se define como el grupo de los veinte países más ricos.

Las regularidades empíricas del mundo capitalista de las últimas cuatro a cinco décadas fueron presentadas en el segundo capítulo. Para comodidad del lector, aquí se les reproduce nuevamente:

(1) Existencia y persistencia del desempleo en el primer mundo.

(2) Existencia y persistencia del desempleo y subempleo en el tercer mundo.

(3) Existencia y persistencia de las diferencias en el nivel de ingreso entre grupos étnicos en el tercer mundo.

(4) Existencia en el corto plazo de correlaciones entre variables nominales —cantidad de dinero, tasa de interés, tipo de cambio— y variables reales —salarios reales, nivel de empleo, nivel de producción—, tanto en el primer mundo como en el tercer mundo.

(5) Existencia de una correlación positiva en el largo plazo entre el salario real y el crecimiento del nivel de la producción, tanto en el primer mundo como en el tercer mundo.

(6) Persistencia en el largo plazo de las diferencias en el nivel de ingreso entre el primer mundo y el tercer mundo —el primer problema de la falta de convergencia en el mundo capitalista—.

(7) Existencia y persistencia en el largo plazo de las diferencias entre el grado de desigualdad en la distribución del ingreso nacional entre el primer mundo y el tercer mundo —el segundo problema de la convergencia—.

Este conjunto de hechos constituye, por otra parte, una descripción estilizada del resultado del proceso económico en el mundo capitalista: la producción y distribución de bienes.

Como se puede inferir de este conjunto de regularidades, a pesar de los progresos en la tecnología, que se reflejan en nuevas máquinas e instrumentos y en innovaciones en la telecomunicaciones; a pesar de los progresos cuantitativos en la producción de bienes de consumo, en mayor cantidad y mayor calidad; y, a pesar de los progresos en la expansión del mercado y la democracia —las dos instituciones básicas del capitalismo—, los problemas sociales del capitalismo son persistentes. Desempleo, subempleo, desigualdades dentro de los países y desigualdades entre países están todos presentes. Este es el mundo social en que vivimos.

¿Cómo se puede explicar el mundo social en que vivimos?

La ciencia económica debe explicar esta realidad. Como se ha mostrado en este libro, «explicar» significa que exista una teoría que puede predecir estos hechos de la realidad. Luego, una buena teoría será aquella que pueda predecir todas estas regularidades empíricas; por contraste, una teoría fallará en explicar si sus predicciones son refutadas por todas o algunas de estas regularidades. Este es el criterio que resulta de la epistemología popperiana.

Sobre las teorías más notables de la actual ciencia económica, llamada la economía estándar, y de acuerdo a la presentación que se hizo en el tercer capítulo, las conclusiones sobre su poder explicativo son las siguientes:

La *teoría neoclásica*, que constituye el paradigma actual, predice los hechos 5 y 6, pero falla en explicar los hechos 1 al 4, y no existen investigaciones empíricas concluyentes sobre el hecho 7.

La *teoría clásica* falla, en explicar los hechos 1 al 5, y no existen investigaciones concluyentes sobre los hechos 6 y 7.

La *teoría keynesiana* predice los hechos 4 y 5, pero falla en explicar los hechos 1 al 3, y no existen investigaciones concluyentes sobre los hechos 6 y 7.

Frente a estos resultados de la economía estándar, ha surgido la necesidad de construir nuevas teorías económicas. Han aparecido en efecto nuevas escuelas y teorías. En este libro se ha presentado una de ellas, la teoría de la economía política de la inclusión-exclusión. Esta teoría parte de un nuevo supuesto general sobre el capitalismo. A diferencia de la economía estándar, supone que existen varios tipos de sociedades capitalistas. Así ya no existe la necesidad de que una sola teoría deba explicar todas las regularidades señaladas.

La teoría de la inclusión-exclusión supone que las distintas condiciones iniciales bajo las cuales las sociedades nacieron al capitalismo

son elementos esenciales de su proceso económico y constituyen, por lo tanto, el criterio para definir los distintos tipos de capitalismo. La historia entra ahora de manera explícita en la teoría económica que busca entender el capitalismo. Según esta teoría, las condiciones iniciales esenciales son dos: la dotación inicial de factores productivos —cantidad de capital por trabajador— y el grado de desigualdad inicial en la distribución individual no solo de los activos económicos, sino también de los activos políticos; asimismo, la teoría supone que en las diferencias en las desigualdades iniciales, el factor esencial es la historia colonial de las sociedades.

La teoría de la inclusión-exclusión se constituye así en un sistema teórico. No será una única teoría que intente explicar el funcionamiento de la sociedad capitalista, como fue el intento de las teorías de la economía estándar. Ahora existirán teorías parciales que buscarán explicar el funcionamiento de cada una de las sociedades que componen el sistema capitalista y luego habrá necesidad de una teoría unificada que buscará explicar el sistema como un todo.

Sobre las teorías parciales se han presentado en el libro tres sociedades abstractas. Para puntualizar que se trata de sociedades abstractas, se les ha dado nombres de letras griegas: épsilon, omega y sigma. Estas sociedades nacieron al capitalismo bajo condiciones iniciales diferentes: épsilon nació socialmente homogénea y no superpoblada; omega nació socialmente homogénea y superpoblada; y sigma nació superpoblada y socialmente heterogénea.

Estas tres sociedades abstractas intentan parecerse, respectivamente, a los tres grupos de países en que se puede dividir el mundo capitalista, según su historia colonial. El primer mundo se compone de países con un legado de haber sido países colonialistas, o con un legado débil de dominación colonial, o sin ninguna historia colonial. Luego está el tercer mundo, que se compone de países superpoblados y con un legado débil de dominación colonial. Finalmente, está el tercer mundo, que se compone de países superpoblados con un legado fuerte de dominación colonial. La significación del legado colonial se refiere a indicadores

como duración del periodo bajo dominación colonial, densidad de la población aborigen y uso de población esclava africana.

Para someterlas a la falsación, se mostraron sendos modelos de las tres teorías parciales (capítulo 4). Los resultados obtenidos se pueden resumir así:

La *sociedad épsilon* intenta explicar el primer mundo. Las regularidades empíricas que tienen que ver con el comportamiento del primer mundo y que entonces desafían a la teoría son los hechos 1, 4 y 5. La teoría, en efecto, predice todos estos hechos.

La *sociedad omega* intenta explicar el tercer mundo, que se compone de países con una débil o ninguna herencia colonial. Las regularidades empíricas que tienen que ver con esta realidad y que desafían a la teoría son los hechos 2, 4 y 5. Esta teoría, en efecto, predice todos estos hechos.

La *sociedad sigma* intenta explicar el tercer mundo que se compone de países con una fuerte herencia colonial, que hizo de ellas sociedades multiétnicas y jerarquizadas en el grado de ciudadanía. Las regularidades empíricas que tienen que ver con esta realidad y que desafían a la teoría son los hechos 2, 3, 4 y 5. Esta teoría, en efecto, predice todos estos hechos.

En suma, los datos de la realidad no refutan las predicciones de los modelos de las tres teorías parciales del capitalismo; por lo tanto, no existe razón alguna para rechazar esas teorías parciales en esta etapa de nuestra investigación. Podemos concluir entonces que tenemos buenas teorías parciales del capitalismo. Parciales porque se refieren a sociedades particulares que conforman el sistema capitalista. Estas teorías explican el funcionamiento del primer mundo y del tercer mundo, tomados separadamente.

¿Cuáles son los factores últimos que explican este mundo social?

Las buenas teorías parciales no conducen necesariamente a una buena teoría unificada. Esto ocurre en la física, donde las buenas teorías del mundo grande —teoría de la relatividad— y del mundo pequeño, subatómico —teoría cuántica—, no son consistentes entre ellas. Ambas no pueden ser válidas. Se necesita en la física una teoría unificada.

El paso de teorías parciales a una teoría unificada exige el cumplimiento de ciertos principios lógicos. En este caso, tales principios se cumplen y, por lo tanto, ha sido posible construir una teoría unificada del capitalismo a partir de las teorías parciales épsilon, omega y sigma.

La siguiente pregunta es si la teoría unificada es una buena teoría, si puede pasar la prueba de la refutación empírica. Para tal efecto, fue necesario construir teorías adicionales sobre el desorden social, el proceso de acumulación de capital físico y de capital humano, que se mostraron en los capítulos del 5 al 7. En el octavo capítulo se construyeron dos modelos de la teoría unificada, una estática y otra dinámica. Las predicciones de estos modelos se confrontaron con los datos de la realidad.

Las regularidades empíricas relevantes para la refutación son los hechos 6 y 7, pues se refieren al funcionamiento del sistema capitalista, tomado como un todo. El resultado es que los modelos de la teoría unificada, en efecto, predicen estos dos hechos. Como los datos de la realidad no refutan las predicciones del modelo estático y del modelo dinámico de la teoría unificada del capitalismo, *no existe razón alguna* para rechazar la teoría unificada en esta etapa de nuestra investigación.

Tenemos así teorías parciales que explican los tipos de sociedades capitalistas, tomados separadamente; en adición, tenemos una teoría unificada que explica el funcionamiento del mundo capitalista, tomado como un todo. La teoría de la inclusión-exclusión parece ser una buena teoría del capitalismo. El nombre de la teoría hace referencia a que el sistema capitalista opera con mecanismos que incluyen a los individuos en

el proceso económico, pero también con mecanismos que los excluyen. Las sociedades más inclusivas son más desarrolladas.

Podemos concluir, entonces, que los países del primer mundo se parecen a la sociedad épsilon; los del tercer mundo con una débil herencia colonial, a la sociedad omega; y los países del tercer mundo con una fuerte herencia colonial, a la sociedad sigma. Los países con una fuerte herencia colonial constituyen la mayoría en el tercer mundo y, por lo tanto, la teoría sigma es la que tiene la mayor relevancia en esta realidad. También debe quedar en claro que la correspondencia entre estas tres sociedades abstractas y los tres grupos de países del mundo capitalista real es una correspondencia estadística; es decir, son válidas en general, en el promedio, aunque existirán excepciones.

El sistema capitalista como un todo puede ser visto como una gran sociedad abstracta que es socialmente heterogénea, pues se compone de las sociedades épsilon, omega y sigma. El mundo capitalista es, por lo tanto, una sociedad de tipo sigma. Según la teoría unificada, sociedades muy desiguales funcionan con desorden social. Así, la teoría unificada también explica por qué existe violencia social a nivel mundial y por qué el grado de violencia social es mayor en el tercer mundo que en el primer mundo.

Hemos construido el concepto de nivel de desarrollo económico de una sociedad como la combinación de dos indicadores: su nivel de ingreso y su grado de desigualdad en la distribución de ingresos. Las regularidades empíricas 6 y 7 implican que los países del primer mundo tienen un mayor nivel de desarrollo económico que los del tercer mundo.

Según la teoría unificada, el factor último que explica estas diferencias en el nivel de desarrollo son las condiciones iniciales de los países —la historia—. Dada esas variables exógenas, la teoría unificada explica por qué las diferencias en el nivel de desarrollo son persistentes; explica por qué no se han producido cambios importantes en estas desigualdades en toda la historia del capitalismo.

El proceso de desarrollo económico lleva a un progreso cuantitativo de los países capitalistas, tales como un mayor nivel de producción, nuevos bienes y nuevas tecnologías. Sin embargo, cualitativamente los países tienden a mantener sus diferencias iniciales. La desigualdad dentro de países y entre países tiende a mantenerse. El resultado teórico fundamental de este libro es que en el funcionamiento del capitalismo la historia cuenta. En el caso particular de los países tipo sigma, el proceso económico no tiene mecanismos para eliminar endógenamente el peso de la historia. Igual conclusión se obtiene para el capitalismo mundial. La paradoja de la falta de convergencia en un mundo capitalista cada vez más integrado queda así resuelta.

¿Por qué no se ha podido cambiar este mundo social?

La conclusión de la teoría unificada, de que el peso de la historia es importante en el funcionamiento del capitalismo, no debe leerse como determinismo histórico. El proceso económico no es igual al proceso físico, en el cual después del inicio del universo, del llamado *Big Bang*, no hay nada que podamos hacer para cambiar su trayectoria.

En el proceso económico existen variables exógenas que pueden cambiarse y de esa manera se pueden cambiar las variables endógenas, es decir, la producción y distribución actual. Pero, a diferencia de lo que sostiene la economía estándar, la teoría unificada establece que las variables exógenas provienen de las condiciones iniciales, es decir de la historia de los países. Cambiar las variables exógenas implica, entonces, romper con la historia.

La utilidad práctica de una buena teoría es que puede servir de fundamento en el diseño de las políticas públicas. Los objetivos de estas pueden basarse en principios normativos o éticos, pero su logro requiere de una teoría positiva sobre el mundo real, pues solo así se podrá asegurar que las acciones que se tomen conducirán al objetivo deseado. Las políticas públicas que se basan en teorías positivas equivocadas no

lograrán los objetivos deseados. El objetivo de la política pública es modificar los resultados del proceso económico, y se puede lograr ese objetivo cambiando el valor de las variables exógenas de una buena teoría, de aquella que ha sido sometida a la prueba de la falsación y no ha fallado.

El diseño de las políticas públicas puede referirse al ámbito nacional o mundial, como se ha mostrado en el noveno capítulo. En base a la teoría unificada, allí se propuso que en el ámbito mundial, el objetivo de la política mundial debería ser la búsqueda del desarrollo económico intergeneracional. La sociedad capitalista mundial debe buscar no solo un nivel de ingreso mayor para nuestra generación, sino también reducir la desigualdad en ingresos dentro de los países del tercer mundo y la desigualdad entre el primer y el tercer mundo. Con «intergeneracional» se quiere decir que la sociedad capitalista mundial debe tomar en cuenta también las necesidades de las generaciones futuras, y no solo las de nuestra generación. Este objetivo implica cuidar del medio ambiente físico.

El instrumento para lograr esos objetivos es modificar una de las variables exógenas de la teoría unificada: la desigualdad en la distribución de activos entre grupos sociales; en particular, la desigualdad en activos políticos. Se debe buscar reducir esta desigualdad, tanto dentro de los países del tercer mundo como entre el primer mundo y el tercer mundo, es decir, tanto a escala nacional como a escala mundial. Según la teoría unificada, una sociedad compuesta de ciudadanos de primera categoría es un mundo mejor para todos, pues las instituciones —incluyendo aquí el mercado y la democracia, las dos instituciones básicas del capitalismo— funcionarán mejor. Para decirlo de otra manera, según la teoría unificada, si la desigualdad en activos se mantiene, la desigualdad de ingresos dentro de los países, así como entre países, también se mantendrá.

Tal vez la redistribución de activos económicos sea socialmente inviable, pero la redistribución de activos políticos parece compatible con un contexto de creciente globalización, democratización y demanda

por derechos, como ocurre actualmente. La igualdad de la ciudadanía a escala mundial sería, en efecto, una forma de romper con la historia.

¿Cómo construir un mundo de ciudadanos homogéneos, todos de primera categoría?, ¿quiénes serían los actores sociales que pudieran llevar a cabo tal tarea y así romper con la historia?, ¿quiénes serían los actores sociales que tienen tanto el poder como los incentivos para llevar a cabo ese cambio en la variable exógena del proceso económico?

Usualmente se responde a esta pregunta mencionando a los gobiernos. Como hemos mostrado en este libro (capítulo 5), los bienes públicos son los que contribuyen a la existencia del Estado y de la sociedad como un todo, y no como la simple suma de individuos. Los bienes públicos introducen la necesidad de la existencia de los gobiernos; también contribuyen a la calidad de vida en la sociedad, pues o todos estarán bien o todos estarán mal, dependiendo de la producción de los bienes públicos. Los gobiernos tienen el poder de redistribuir activos y generar así mayor orden social, pero no tienen los incentivos para llevar a cabo estas acciones. Intentar la redistribución de activos no les servirá de mucho para ganar las próximas elecciones, que es su motivación principal.

En realidad ningún agente social que hemos estudiado puede hacerlo. Los capitalistas también tienen el poder pero no tienen el incentivo. Ellos, pagan un costo por el desorden social, que se origina en la desigualdad, pues sus ganancias serían mayores en un mundo de ciudadanos de primera categoría. El mayor orden social los beneficiaría. Pero el orden social es un bien público y ellos no tienen los incentivos para producirlo. Con la motivación del interés propio, no se puede producir un bien público porque el problema de todos se convierte en un problema de nadie.

Los trabajadores tienen el incentivo pero no tienen el poder. Para tener poder y producir los bienes públicos que les interesa, como los derechos, necesitan llevar a cabo acciones colectivas. Pero las acciones colectivas enfrentan limitaciones. Está el problema olsoniano, que tampoco es ajeno al grupo de trabajadores. Existen dos limitaciones

posiblemente más importantes en los trabajadores del tercer mundo. Una de ellas es la pobreza. Las acciones colectivas implican costos y los ingresos de los trabajadores pueden no ser suficientes como para autofinanciar la acción colectiva. La otra es la exclusión del derecho ciudadano. Ciudadanos de segunda categoría no pueden tener voz; por lo tanto, el costo de la acción colectiva es muy alto y el beneficio muy incierto.

La producción y distribución en el mundo capitalista es una situación de equilibrio. Por lo tanto, ningún actor social cuenta con el poder y el incentivo para cambiar esta situación.

Llegamos al punto en que la solución del problema económico está fuera del proceso económico. Llegamos así al problema de Gödel en la ciencia económica: el proceso económico no puede comprenderse así mismo, ni tampoco puede cambiarse así mismo. Si todo proceso económico pudiese comprenderse así mismo; y si todo problema económico pudiese resolverse dentro del proceso económico, es decir, endógenamente, no existirían problemas sociales por resolver. Todos los problemas sociales se autorregularían. La existencia y persistencia de un mundo muy desigual y que opera con desorden social, que son las regularidades básicas del capitalismo, nos indican que ese mecanismo no existe.

La solución al problema económico de la desigualdad global vendrá de procesos distintos al económico, de procesos nuevos que producirán los actores sociales que se requieren. Como toda innovación, esto tomará tiempo. Tal vez este proceso se refuerce y se agilice con la mayor conciencia que tenga la gente sobre el mundo social en que vivimos. Un ejemplo claro y alentador de este efecto es el del medio ambiente, que se ha convertido en un problema global y muy serio, gracias a la literatura científica que nos ha alertado sobre el problema del calentamiento global.

Este libro, amigo lector, a través de la mejor comprensión que usted tenga ahora sobre el mundo social en que vivimos, puede contribuir a desarrollar esa conciencia y ese nuevo proceso de cambio.

Apéndice

Economía y las ciencias naturales

Las ciencias naturales estudian las relaciones entre los objetos materiales del mundo, mientras que las ciencias sociales estudian las relaciones entre las personas en una sociedad dada. Además del campo de estudio, ¿cuál es la principal diferencia entre las ciencias naturales y sociales? Con el fin de responder esta gran pregunta, en primer lugar tomaremos el caso particular de tres disciplinas: economía, física y biología. La física estudia materiales inanimados, mientras que la biología estudia los organismos vivientes.

Muchas personas han argumentado que la economía no es en absoluto una ciencia. El contenido de este libro los habrá persuadido de que la economía sí es una ciencia. ¿Pero qué tipo de ciencia es la economía? Usualmente la física es vista como la ciencia ejemplar, el paradigma de las ciencias. La biología es una ciencia, pero no como la física. ¿Es la economía una ciencia como la física o como la biología?, ¿tenemos ciencias distintas o una familia de ciencias con una misma epistemología general? En este último caso, ¿qué es lo común a todas las ciencias? Este apéndice intenta responder estas preguntas.

Sobre la epistemología

¿Son los métodos de investigación distintos entre la economía y las ciencias naturales? ¿Cuáles son las alternativas al método popperiano que se ha utilizado en la economía en este libro?

El conocimiento científico puede verse como un bien económico, el cual es útil pero escaso. El conocimiento es un bien mental, un entendimiento mental de la realidad; es un bien escaso y debe producirse. ¿Cómo se produce el conocimiento científico? Como ocurre con las personas, el conocimiento científico camina en dos pies: el conocimiento sobre los hechos de la realidad y las razones detrás de esos hechos, la teoría. En esencia, el conocimiento científico se logra que exista siempre consistencia entre hechos y teoría.

El conocimiento científico significa conocimiento libre de errores. ¿Cuáles son las fuentes posibles de error? Podrían ser:

(a) la medición de los hechos de la realidad;
(b) la lógica interna del sistema teórico utilizado para explicar esos hechos;
(c) la lógica que conecta teorías y hechos.

El conocimiento científico es más riguroso cuando estos errores son limitados o minimizados. La metodología es la tecnología para producir el conocimiento científico. Como cualquier *tecnología*, incluye el conjunto de *métodos* eficaces para producir el conocimiento. Es común distinguir en la literatura sobre metodología de las ciencias tres métodos principales: inductivismo, deductivismo y falsación (Achistein 2004).

El método de inducción busca producir el conocimiento tomando una observación particular de la realidad y haciendo las inferencias universales desde esos hechos. Este método puede llevarnos a un error significativo. Como el filósofo de ciencia Karl Popper dijo, «no importa cuántos cisnes blancos hayamos observado, esto no justifica la conclusión que todos los cisnes son blancos» (Popper 1959: 27).

El problema lógico con el inductivismo está en la pregunta, ¿cuál es el principio que permite justificar la inferencia? Consideremos el siguiente ejemplo: «En cientos de observaciones, los cisnes eran blancos; entonces todos los cisnes son blancos». ¿Por qué?, ¿cuál es la justificación de esta inferencia? «En cientos de observaciones, los cisnes eran blancos porque ellos comieron X; entonces todos los cisnes son blancos

porque ellos comen X». ¿Por qué? «En cientos de observaciones, los cisnes eran blancos porque ellos comieron X que generó las células Y; entonces todos los cisnes son blancos porque ellos comen X que generan las células Y». ¿Por qué? Y así sucesivamente, entramos al problema lógico de la regresión infinita.

Como Popper ha señalado, este ejemplo ilustra que el principio de inducción es otra inducción. La única manera de justificar una inferencia inductiva es a través del uso de un principio que tendrá que ser una afirmación universal. Ahora, si intentamos comprender la realidad a partir de la experiencia, entonces para justificar este método de investigación nosotros tendríamos que utilizar la inferencia inductiva, la cual puede justificarse por una inferencia de orden superior, y así sucesivamente. Popper concluyó, «Así el esfuerzo por basar el principio de inducción en la experiencia previa se rompe, pues nos lleva a una regresión infinita» (Popper 1959: 29). Esto es, precisamente, lo que muestra el ejemplo anterior.

Si el principio de la inducción es otra inducción, lo cual nos lleva al retroceso infinito, entonces no puede existir una lógica inductiva y, por consiguiente, el inductivismo no puede constituir un método para el conocimiento científico. La filósofa Susan Haack ha sentenciado que la «Inducción es injustificable; no puede existir la lógica inductiva» (2003: 35).

Otro problema del método inductivo es la distinción lógica entre asociación y causalidad. La inferencia inductiva usa asociaciones empíricas o correlaciones. Las razones bajo esas asociaciones —la pregunta científica del por qué— no pueden resolverse por el método de inducción, como se ha mostrado en el ejemplo anterior. Esas razones solo pueden ser teóricas, de modo que podemos escaparnos de la trampa del retroceso infinito del inductivismo. La correlación no puede implicar causalidad.

El otro método es el deductivismo, el cual busca producir conocimiento únicamente a partir de los pensamientos. La teoría no solo tiene sus propios supuestos de manera explícita, sino que también

introduce otro supuesto implícito: la teoría lógicamente correcta es empíricamente verdadera. Esta conclusión no puede justificarse. En las ciencias fácticas, por lo tanto, el riesgo de error del método deductivo es enorme.

El tercer método es el de la falsación. Este método se basa en las interacciones entre teoría y realidad. Popper desarrolló este método en su famosa obra *La lógica del descubrimiento científico* (1934). La falsación es un método que reduce el error de aceptar las teorías que son falsas, tales como la aceptación de proposiciones tautológicas o de proposiciones que dependen solamente del aparente «realismo» de sus supuestos.

La falsación no es, sin embargo, un método que sirva para descubrir una teoría que sea verdadera. Si las predicciones de una teoría son refutadas por los datos de la realidad, la teoría es claramente falsa. Si las predicciones de una teoría coinciden con los datos empíricos, ¿se podría decir que la teoría es verdadera? Popper rechaza esta conclusión porque nos estaríamos basando en el método inductivo: de la verdad de casos particulares se estaría haciendo la inferencia de que la teoría es verdadera. «No se puede mostrar la validez empírica de una teoría a partir del razonamiento deductivo. Pero tampoco puede mostrarse esa validez a partir de sus éxitos pasados puesto que estaríamos usando la inducción para justificarse» (Achinstein 2004: 132). ¿Cuántos casos exitosos deben observarse para concluir que una teoría es verdadera? No existe respuesta, tal como lo muestra el ejemplo de los cisnes.

En esencia, no se puede probar que una teoría sea verdadera; solo se puede decir que los datos de la realidad han corroborado las predicciones de la teoría. Todo lo que podemos decir es que la teoría es consistente con la realidad. En el método de la falsación, «falso» no tiene por oposición a «verdadero», sino a «consistente».

¿Es la falsación el método más eficiente? Algunos filósofos de la ciencia, como Thomas Kuhn y Paul Feyerabend han negado la existencia de un método universal para las ciencias (Achistein 2004). Sin embargo, comparado al inductivismo y al deductivismo, la falsación es el método que minimiza los errores que puedan ocurrir en el proceso

de producción del conocimiento. La falsación es, por consiguiente, el método más eficiente comparado al inductivismo y al deductivismo.

La metodología es, así, un conjunto de métodos para producir conocimiento científico. Cada método constituye un sistema lógico del cual se derivan reglas para producir conocimiento; cada método constituye, entonces, una teoría normativa de la epistemología. Esta teoría normativa es una teoría del conocimiento, la teoría de teorías, es decir, una metateoría. Popper propuso la teoría de falsación. Esta es la teoría que se ha aplicado para el estudio de la economía. Para hacerla operativa, se ha construido un modelo particular de esta teoría, llamado alfa-beta. ¿Puede también aplicarse el método alfa-beta a la física y biología?

Sobre la física

Recientemente, los físicos han escrito libros para popularizar el nivel de conocimiento obtenido en esta ciencia. El físico más popular es el británico Stephen Hawking. Basado en su libro *Una breve historia del tiempo* (1996), se presentará aquí la estructura de la física en términos de las proposiciones teóricas y empíricas.

Las primeras proposiciones hechas en la física eran hipótesis empíricas que no estaban asociadas a ninguna teoría. Este tipo de hipótesis, desprovisto de una teoría, se denominará aquí *hipótesis H.* Dos de las más importantes son:

H (1): La Tierra es el centro del universo (Plotomeo, I.A.D).

H (2): El Sol es el centro del universo, y la tierra y los otros planetas se mueven en orbitas circulares a su alrededor. El Sol es estacionario (Copérnico, *c.* 1514).

El descubrimiento del telescopio (1609) fue suficiente para refutar ambas hipótesis. Dos observaciones empíricas surgieron con la ayuda de este instrumento. Este tipo de observaciones empíricas puede llamarse la «observación O». En este tipo de observación empírica no se muestra

ninguna correlación estadística entre variables. Las observaciones que refutaron las hipótesis señaladas anteriormente fueron:

H (1) ≠ O (1): No todos los cuerpos del espacio giran alrededor de la Tierra (Galileo).

H (2) ≠ O (2): Los planetas se mueven en órbitas elípticas alrededor del Sol (Kepler).

La teoría más famosa de la física es la teoría de gravedad propuesta por Newton en 1687. Se le puede enunciar así:

ϕ_1: **Teoría de gravedad.** El universo es un sistema estático en el cual los cuerpos son atraídos entre sí por una fuerza, la cual es más fuerte cuanto más pesados sean los cuerpos y cuanto más cercanos estén uno del otro.

Las predicciones empíricas pueden ser derivadas lógicamente de la teoría ϕ_1. Estas proposiciones serán llamadas γ_1, es decir ϕ_1 implica γ_1. Ellas son:

γ_1 (1): los planetas se mueven en órbitas elípticas alrededor del sol.

γ_1 (2): la luz es una partícula y, como en el caso de una bola de cañón, puede ser agarrada si el observador corre lo suficientemente rápido.

γ_1 (3): la atracción entre objetos opera instantáneamente, tanto sobre distancias cortas como largas.

¿Son estas hipótesis empíricas consistentes con la realidad? Las observaciones empíricas siguientes pueden servir para buscar la consistencia con esta teoría:

γ_1 (1) = O (2). La observación empírica de Kepler corrobora la primera predicción empírica.

γ_1 (2) ≠ O (3): Nada puede viajar más rápido que la luz (Hawking 1996: 32). Los físicos experimentales han mostrado que la luz viaja a 670 millones de millas por hora (es decir, 300 mil kilómetros por segundo), sin tener en cuenta la velocidad del observador.

γ_1 (3) ≠ O (3).

Un caso de inconsistencia empírica es suficiente para rechazar una teoría; por consiguiente, los datos empíricos refutan la teoría de la gravedad. La pregunta ahora consiste en encontrar otra teoría que prediga la proposición γ_1 (1), pero no las otras.

Albert Einstein desarrolló dos teorías a inicios de 1900. La primera es la Teoría de Relatividad. Esta teoría y sus predicciones empíricas pueden enunciarse como sigue:

ϕ_2: **La Teoría especial de la relatividad.** Tiempo y espacio no son categorías absolutas, sino que dependen del movimiento relativo de los observadores.

γ_2 (1): ningún objeto puede viajar más rápido que la luz.

γ_2 (2): la relatividad comienza a ser más significativa cuando los observadores alteran sus velocidades acercándose a la velocidad de la luz. Estando sobre la tierra, todos los observadores ven el mismo movimiento de objetos porque las diferencias de velocidad entre los observadores son demasiado pequeñas. Si un observador estuviera en el espacio viajando a velocidades cercanas a la velocidad de la luz, podría observar el movimiento de objetos de una manera diferente comparada a lo que vería un observador desde la Tierra.

γ_2 (3): la velocidad de la luz parecer ser la misma para todos los observadores (Hawking 1196: 39).

γ_2 (4): $E=mc^2$, donde E es la energía, m es la masa y c la velocidad de la luz. Esta ecuación es considerada la más famosa de la ciencia.

El supuesto implícito de la Teoría de gravedad es que el tiempo es absoluto. En la teoría de relatividad especial el supuesto es que cada observador tiene su propia medida del tiempo, como si estuviera grabada en un reloj que cada uno llevara dentro; es decir, los relojes llevados por observadores diferentes no estarían de acuerdo si ellos estuvieran moviéndose a velocidades diferentes. La velocidad de luz, sin embargo, es la misma para cada observador, sin importar cuán rápido este se esté moviendo.

La consistencia empírica de teoría de relatividad puede mostrarse como sigue:

γ_2 (1) = O (3). Se pueden mencionar ejemplos que muestran que la luz viaja más rápido que cualquier otro objeto. En una tormenta, uno ve la luminosidad antes de oír el trueno. En un estadio de béisbol, cuando un jugador pega la pelota, uno ve la pelota golpeada antes de oír el sonido.

γ_2 (2) $\approx$ O (sin datos)

2 (3) = O (4): el experimento de Michelson-Morley (Hawking 1996: 39).

2 (4) = O (5): la explosión de bombas atómicas.

En suma, los hechos disponibles no refutan la teoría especial de la relatividad.

La Teoría de la gravedad y la Teoría especial de la relatividad están en conflicto respecto a la velocidad de luz; es decir, ambas no pueden ser verdaderas. «[La teoría de la gravedad] sostiene que los objetos son atraídos unos a otros con una fuerza que depende de la distancia entre ellos. Esto significa que si uno moviera uno de los objetos, la fuerza de atracción sobre el otro cambiaría instantáneamente. O en otros términos, los efectos

gravitacionales deben viajar con velocidad infinita, y no por debajo de la velocidad de luz o igual a ella» (Hawking 1996: 39).[9]

Así, queda claro que γ_2 (1) $\neq \gamma_1$ (2). La observación empírica O (3), que nada viaja más rápidamente que la luz, refuta la teoría de la gravedad, pero no a la teoría de la relatividad especial. Pero, entonces, ¿cómo queda la fuerza de gravedad entre los cuerpos?

La segunda teoría de Einstein es la Teoría de la relatividad general, que junto con sus predicciones empíricas, puede enunciarse como sigue:

> ϕ_3: **La Teoría general de la relatividad.** La gravedad es el resultado de las distorsiones en la geometría del tiempo-espacio. [10] La gravedad no es una fuerza similar a otras fuerzas; es, más bien, una consecuencia del hecho que el espacio-tiempo no es plano —como había sido previamente supuesto— sino que es curvo, debido a la distribución de la masa y la energía. Los cuerpos siempre siguen líneas rectas en el espacio-tiempo de cuatro dimensiones, pero ante nosotros, en nuestro espacio tridimensional, parecen moverse en senderos curvos (Hawking 1996: 40).
>
> γ_3 (1): los planetas siguen órbitas elípticas alrededor del sol (1996: 40).
>
> γ_3 (2): la luz debería ser doblada por la fuerza de la gravedad. Por ejemplo, los conos de luces cercanos al sol podrían ser ligeramente torcidos hacia adentro, debido a la masa solar (Hawking 1996: 42).

---

[9] Otro ejemplo: de acuerdo con la Teoría de la gravedad, si el sol explotara, la Tierra podría sufrir instantáneamente una alteración, saliendo de su órbita elíptica. Pero de acuerdo con la Teoría general de la relatividad, este efecto no sucedería, debido a que no hay información que pueda ser transmitida más rápido que la velocidad de la luz; por esto, este efecto, de la caída de la Tierra sería instantáneo. Esto podría tomar ocho minutos, que es el tiempo que demora la luz en llegar del sol a la Tierra.

[10] Espacio - tiempo es la descripción dimensional del universo, que junta las tres dimensiones del espacio y la única dimensión del tiempo.

γ_3 (3): El tiempo debería transcurrir más lentamente cerca de un cuerpo macizo como la tierra (1996: 43)

γ3 (4): El universo tuvo un tiempo inicial, la del *Big Bang* (Hawking 1996: 44).

Sobre la refutación empírica de estas predicciones, los datos son los siguientes:

γ_3 (1) = O (6): Las órbitas medidas por un radar están de acuerdo con la teoría. «De hecho, las órbitas predichas por la teoría de la relatividad general son casi exactamente iguales que aquellas predichas por la teoría de la gravedad de Newton [$\gamma_1(1) = \gamma_3(1)$]. Sin embargo, en el caso de Mercurio, por ser el planeta más cercano al sol, el efecto de la gravedad es mayor y las predicciones de la teoría de la relatividad general muestran una pequeña desviación con respecto a las predicciones de Newton. Este efecto fue notado antes de 1915 y sirvió como una de las confirmaciones de la teoría de Einstein» (Hawking 1996: 40-42).

γ_3 (2) = O (7): Al observar un eclipse desde África oriental en 1919, una expedición británica mostró que la luz fue desviada por el sol. Esta desviación de la luz ha sido confirmada por varias observaciones posteriores (Hawking 1996: 42).

γ_3 (3) = O (8): Los experimentos han corroborado esta predicción (1996: 43).

γ_3 (4) ≈ O (sin datos).

En suma, las observaciones empíricas disponibles no refutan las predicciones de la Teoría general de la relatividad.

La Teoría general de la relatividad estudia los cuerpos grandes del universo. Los fenómenos en cuerpos sumamente pequeños son estudiados por la teoría de la mecánica cuántica. Hoy día contamos con «microscopios» de alta calidad, que son distintos a los microscopios

regulares porque, de una parte, son instrumentos que nos permiten ver moléculas, átomos y pequeñas partículas, y porque algunos de ellas son instrumentos muy grandes. Estos instrumentos han hecho posible un mayor conocimiento en las observaciones empíricas del mundo subatómico.

Durante mucho tiempo, se supuso que la materia estaba compuesta de átomos. El átomo era el último elemento inobservable y las teorías de la física hacían supuestos sobre su composición y comportamiento. A inicios de la década de 1930, los átomos llegaron a ser observables. Ahora, nosotros sabemos que los átomos están constituidos por un núcleo —conteniendo protones y neutrones— y por electrones. Pero, ¿de qué están hechos estos elementos? Como no eran observables, se hicieron supuestos sobre sus composiciones y comportamientos. En 1968, se llegó a observar que el núcleo estaba hecho de elementos aun más pequeños, llamados *quarks*. Surge de nuevo la pregunta, ¿de qué están hechos los *quarks* y los electrones? Hoy, algunos físicos suponen que el último elemento inobservable es un elemento llamado la cuerda —la teoría de las cuerdas—.

La teoría cuántica y sus predicciones empíricas pueden ser presentadas de la siguiente manera:

> ϕ_4: **La teoría cuántica.** Las partículas no tienen ni posiciones ni velocidades exactamente determinadas. Cuando se le examina en distancias cada vez más pequeñas y en escalas de tiempo cada vez más cortas, el universo es un lugar incierto, gobernado por el azar. Esta teoría se refiere al mundo subatómico.
>
> > γ4 (1): Las observaciones sobre la posición y velocidad de objetos en el mundo subatómico son inciertas. Este es el *principio de incertidumbre* de Heisenberg, establecido en 1926. Este principio dice que «uno nunca puede estar totalmente seguro ni de la posición ni de la velocidad de una partícula; el conocimiento más preciso de uno implica el conocimiento menos preciso del otro» (Hawking 1996: 243).

La teoría de la mecánica cuántica no predice un solo resultado para una observación, sino un número de posibles resultados, y nos dice cuán probable es cada uno de estos; por consiguiente, introduce un elemento inevitable de incertidumbre en la física (1996: 73). Sobre su consistencia empírica, se puede decir lo siguiente:

> γ_4 (1) = O (9): esta concuerda perfectamente con los experimentos (1996: 73).

Según la Teoría general de la relatividad, el universo es un lugar liso o llano en la geometría espacial en escalas macroscópicas; sin embargo, según la teoría de la mecánica cuántica, es una arena caótica en escalas microscópicas. ¿Pueden ser estas dos teorías correctas? Si todo en el universo depende de todo lo demás —un requisito de la teoría unificada—, la respuesta es no.

Hoy los científicos describen el universo en términos de dos teorías parciales básicas: la Teoría general de la relatividad y la mecánica cuántica. Lamentablemente, estas dos teorías son conocidas por su inconsistencia de una con la otra, ambas no pueden ser correctas (Hawking 1996: 18).

La geometría espacial lisa de la relatividad general se rompe en las escalas de distancias cortas debido a las violentas fluctuaciones de las partículas de la mecánica cuántica. Empíricamente, la realidad no es consistente con la teoría unificada que integra relatividad general y la mecánica cuántica. Esta teoría unificada predice que ciertas cantidades, como la curvatura del espacio-tiempo, son infinitas, pero estas cantidades cuando son medidas pueden ser perfectamente finitas (Hawking 1996: 215).

¿Cómo podría ocurrir esto?, ¿cómo puede existir orden a grandes escalas pero desorden a escalas pequeñas?, ¿puede generarse orden a partir del desorden?, ¿cómo podría resolverse esta contradicción? Se necesita una teoría unificada de las fuerzas de la naturaleza, fuerzas que sean independientes de la escala de los objetos, desde las pequeñas distancias dentro del mundo subatómico a las distancias más grandes en el

inmenso universo. Esta es la pregunta más desafiante en la física de hoy: encontrar una teoría unificada de la física, *la teoría del todo.*

Como hemos visto, las teorías de la física (ϕ) dan lugar a proposiciones que son empíricamente refutables (γ). Hemos utilizado un método de falsación que no es el alfa-beta, que podríamos llamar el *método directo de falsación* dentro de la epistemología popperiana. Este método de falsación puede ser utilizado en la construcción del conocimiento en la física. Es *como si* los físicos hubieran usado la epistemología de la falsación en el desarrollo de la física. El uso de la falsación en la física muestra que el conocimiento ha avanzado enterrando teorías, de funeral en funeral.

El método alfa-beta no es aplicable a la física. La razón es simple: aunque las relaciones entre los objetos materiales se repiten periodo tras periodo, ellas no constituyen un proceso (en el sentido del término definido en capítulo 1 y sintetizado en la figura 1.1). No existe ninguna variable exógena en el proceso físico. No existe nada fuera del universo físico que pueda cambiar independientemente —o exógenamente—, que ingrese al sistema, y pueda mover el universo de una situación de equilibrio —estático o dinámico— a otra. Las relaciones físicas se presentan como *leyes* porque no existen variables exógenas que puedan modificar esas relaciones.

El conjunto de proposiciones (α,β) usado en economía tiene diferente contenido comparado al conjunto de proposiciones (ϕ,γ) que se ha utilizado aquí para la física. Las proposiciones beta están referidas a los efectos de las variables exógenas sobre las variables endógenas; pero este no es el caso en las proposiciones gamma. Incluso la naturaleza de falsación empírica es diferente en la física. En la física, la expresión $\gamma \approx O$ se refiere a la pregunta de consistencia entre la predicción de la teoría con un conjunto de observaciones empíricas, las cuales son obtenidas con la ayuda de instrumentos de medida, tales microscopios y telescopios; por contraste, en la economía, $\beta \approx b$ se refiere a la consistencia entre las predicciones de la teoría y el conjunto de datos que representan la asociación o correlación estadística entre variables

endógenas y exógenas. Debido a la existencia de variables exógenas, la economía es una ciencia, pero no es como la física. La economía es una ciencia diferente, más compleja.

Sobre la biología evolutiva

Los biólogos también han escrito recientemente libros para un gran auditorio, como Mayr (1997), Casti (2001), Smith (2002) y Pasternak (2003). Estos trabajos se usarán aquí para presentar la estructura de esta disciplina.

Según Ernst Mayr, biólogo de Harvard, la biología es, como la física, una ciencia; pero la biología no es una ciencia similar a la física. Existen dos campos en biología que son bastante diferentes en la naturaleza de conocimiento. Una es la biología molecular, que es más parecida a la física, y la otra es la biología evolutiva, que estudia las poblaciones. Se utilizará este segundo campo para hacer la comparación con la economía.

La biología evolutiva puede representarse como un proceso porque en el agregado hay la posibilidad de repeticiones y porque en ella existen variables exógenas que pueden cambiar el equilibrio en que viven los seres en el mundo. Esta variable exógena es el ambiente físico. La biología evolutiva estudia las interacciones entre los numerosos organismos, cada uno de ellos sumamente complejos.

¿Pueden construirse modelos teóricos para explicar estas interacciones complejas? La respuesta es afirmativa. En efecto, en biología se utilizan modelos matemáticos.

Según los biólogos, la razón es que el uso de ecuaciones permite predecir la conducta de los organismos vivientes. Esta predicción es, sin embargo, usualmente solo cualitativa. La exacta consistencia numérica con los datos es difícil y es demasiado esperar porque en cualquier modelo se utiliza la abstracción.

¿Cuál es la justificación para omitir algo que ciertamente afecta el resultado en un modelo? Los biólogos lo justifican así. Primero, si fuera importante, el modelo no dará las predicciones correctas, incluso cualitativas. Segundo, si nosotros intentamos poner todo en un modelo, este será simplemente inútil. En la biología, solo los modelos simples son útiles. El precio que los biólogos deben pagar por esta simplificación es la falta de exactitud cuantitativa entre sus predicciones y los datos (Smith 2002). Nótese que esta es la misma epistemología que utilizamos en la economía.

La teoría darwiniana puede reformularse en lo que se refiere a teorías parciales usando el método alfa-beta, como sigue:

> α_1: **La Teoría del descendiente común.** Los organismos vivientes han evolucionado de una raíz común. Por lo que se refiere a su origen, los organismos vivientes constituyen un solo árbol familiar.
>
> > $\beta_1(1)$ Animales y plantas de recientes períodos geológicos serán descendientes de aquellos de los períodos geológicos más antiguos. Debemos observar una secuencia histórica en la existencia de organismos vivientes.
> >
> > $\beta_1(1)$= O (1): La secuencia observada de los fósiles que se han encontrado en los estratos de la tierra es consistente con la predicción. La teoría habría sido falsa si estos fósiles de elefantes y jirafas hubiesen sido encontrados en el período del Cretáceo Temprano.
>
> α_2: **La teoría de la selección natural.** En un ambiente dado, algunos tipos de individuos tienen mayor probabilidad de sobrevivir y reproducirse que otros.

Los supuestos de esta teoría incluyen un proceso con los siguientes aspectos: (a) los descendientes deben parecerse a sus padres; (b) existe variación genética en los organismos vivientes; (c) los cambios en la composición genética de los organismos vivientes es aleatoria, pero la

eliminación de individuos no es aleatoria. La teoría no establece qué tipo de individuos sobrevivirá y por qué; solo predice que algunos individuos podrán sobrevivir.

El equilibrio del sistema es dinámico. Dado el medio ambiente, hay un equilibrio en la distribución aleatoria de las características de los individuos. Cuando estos individuos se reproducen, ellos transfieren sus características a su descendencia, cuyo resultado es una población de individuos con características apropiadas para la supervivencia. Las características del equilibrio presentes en la población de una generación es una muestra aleatoria de aquellas que están presentes en la generación anterior, y así sucesivamente. El supuesto implícito es que las características innatas de individuos son heredadas, pero las características adquiridas no lo son.

Las predicciones empíricas que pueden derivarse de esta teoría incluyen lo siguiente:

> $\beta_2(1)$: **Hipótesis de la supervivencia del más apto.** Los cambios en el ambiente causarán una eliminación no aleatoria de algunos individuos. La supervivencia del más apto en la competencia por el recurso escaso en el ambiente prevalecerá.
>
> $\beta_2(2)$: **Hipótesis de la evolución.** Los organismos vivientes no son estáticos; ellos evolucionan guiados por la selección natural en el tiempo.
>
> $\beta_2(3)$: **Hipótesis de la multiplicación de las especies.** Las mismas especies que viven en ambientes diferentes se convertirán en especies distintas. En otras palabras, las mismas especies en ambientes diferentes mostrarán rasgos distintos, entendidos como diferencias en promedio y varianza. La separación por montañas, mares o la discontinuidad ecológica generará este resultado.

La evidencia empírica parece no refutar estas predicciones.

En suma, las relaciones entre organismos vivientes pueden ser representadas por un proceso: el proceso biológico. Las condiciones para este

reduccionismo se cumplen plenamente: existe repetición en el equilibrio de estas relaciones y existen variables exógenas que modificarán la situación de equilibrio de estas relaciones. Por consiguiente, el método alfa-beta puede aplicarse al estudio de biología. Epistemológicamente, la economía se parece más a la biología que a la física.

Sobre el universalismo ontológico

La próxima comparación entre la economía y las ciencias naturales se refiere a la pregunta sobre el universalismo ontológico, el supuesto de un único mundo. La economía considera a la economía mundial como su «universo». En la física y la biología, el universo relevante y comparable es la Tierra.

Las discrepancias cuantitativas entre las teorías de Einstein y la teoría de Newton son sumamente pequeñas en nuestro planeta. Si uno lanza una bola de béisbol, las teorías de Newton y de Einstein podrán ser usadas para predecir donde aterrizará, y las respuestas serán diferentes, pero la diferencia será pequeña, tan pequeña que ellas estarán más allá de nuestra capacidad para detectar experimentalmente esa diferencia (Greene 2003: 76). La teoría de Newton no es válida para todo espacio y tiempo, pero la teoría de Einstein sí lo es.

En el mundo limitado del planeta Tierra, sin embargo, tanto la teoría de Newton como la teoría de Einstein son consistentes con las observaciones empíricas. Tenemos aquí un ejemplo de una realidad con dos teorías. También tenemos un ejemplo de una teoría refutada que, sin embargo, todavía puede ser útil, pues algunas de sus predicciones son empíricamente consistentes. La refutación de una teoría no implica su rechazo total. El conocimiento es acumulativo y todas las teorías, buenas o malas, juegan un papel en este proceso. Considerando solo el plantea Tierra como realidad, podríamos decir que la teoría de Newton es más fina y simple que la de Einstein.

Si nuestro planeta pudiera separarse en segmentos, utilizando cualquier tipo de criterio, estos segmentos no serían diferentes en el sentido que la Teoría de la gravedad podría explicar las relaciones entre los objetos, no importa cómo esos segmentos fueron creados. Los átomos se comportarán exactamente de la misma manera en cualquier parte del planeta. La Teoría de la gravedad es una teoría general para el planeta. El principio de universalismo ontológico se aplica entonces en la física.

En la biología evolutiva, el universo puede separarse en segmentos diferentes utilizando como criterio el ambiente físico, que es la principal variable exógena. En cada ambiente, así definido, los organismos vivientes se comportarán de modo diferente. Por ejemplo, las plantas se comportan de manera diferente en los trópicos que en las zonas templadas, en las altas montañas que en las costas. La diversidad de realidades biológicas implica la construcción de teorías parciales que puedan explicar la fisiología vegetal en cada una de esas realidades específicas y particulares. La pregunta sobre la unidad del conocimiento aparece en la biología.

Las plantas se comportan de modo diferente en los Andes respecto a las llanuras americanas, ya que su fisiología es diferente. En las zonas templadas, las plantas adaptaron su fisiología a ambientes en los cuales las mayores variaciones en la temperatura ocurren en las estaciones a lo largo del año. En los Andes, en cambio, las plantas han tenido que adaptarse a un ambiente en el que las mayores variaciones de temperatura se dan durante el día, y no durante las estaciones. En ambas zonas, sin embargo, la teoría de la fotosíntesis es la misma. La fotosíntesis es la teoría general que lleva a aceptar la existencia de una teoría unificada de la fisiología vegetal.

El proceso económico también tiene lugar en ambientes sociales diferentes. Como se ha mostrado en este libro, el sistema capitalista mundial puede ser dividido en varios tipos de sociedades capitalistas usando el criterio de las condiciones iniciales —la historia—, que incluye las dotaciones de factores iniciales y el grado inicial de desigualdad. En cada ambiente social las personas se comportan de manera diferente y la sociedad funciona de manera distinta. La unidad del conocimiento

es entonces un problema importante a resolver. Este libro ha intentado proporcionar una solución. Épsilon, omega, y sigma son teorías parciales que han podido ser consolidadas en una teoría unificada de capitalismo.

La ontología universalista —el supuesto de un único mundo real— es aplicable a la física, pero no a biología ni a la economía. La economía es, nuevamente, una ciencia que se parece más a la biología que la física.

Sobre la cuestión de la medición

«Ciencia es medición», reza un conocido principio científico. Sin embargo, la cuestión más desafiante consiste en poder distinguir entre las relaciones que son medibles de aquellas que no lo son. Por ejemplo, ¿existe un mundo real, observable, independiente de nuestro pensamiento? Ciertamente, se necesita un criterio para determinar lo que es y no es medible.

El filósofo John Searle (1995) ha propuesto un criterio. Él señala que las proposiciones sobre los hechos pueden ser clasificadas como ontológicas y cognitivas; además, cada categoría puede a su vez ser dividida entre objetiva y subjetiva. Por lo tanto, se puede construir una matriz dos por dos, tal como se muestra en la tabla A.1.

Una proposición sobre la realidad es cognitivamente subjetiva cuando su verdad o falsedad no es una cuestión factual simplemente, sino que depende del punto de vista de la persona; es cognitivamente objetiva cuando la verdad o su falsedad es independiente del punto de vista de la persona. La primera es una proposición normativa y la segunda es una proposición positiva. De otro lado, una proposición sobre la realidad es ontológicamente subjetiva cuando su existencia depende de los sentimientos de los individuos; es ontológicamente objetiva cuando su existencia es independiente de cualquier sentimiento. La primera se refiere a una situación mental y la segunda se refiere a un objeto físico.

Tabla A.1. Tipos de realidad basados en el criterio de clasificación de Searl

	Cognitivo	
Ontológico (existencia)	Objetivo (positivo)	Subjetivo (normativo)
Objetivo (material)	Este papel es delgado	Odio el papel delgado
Subjetivo (mental)	La cantidad de dinero aumentó	Adoro el dinero

Toda proposición sobre la realidad puede ponerse en una celda de la matriz y tendrá dos sentidos, uno ontológico y otro cognitivo. Considere los ejemplos mostrados en la tabla A.1. Las mediciones de un pedazo de papel son ontológicamente objetivas y cognitivamente objetivas. Los sentimientos sobre un pedazo de papel serán antológicamente objetivos y cognitivamente subjetivos. Las mediciones de la cantidad de dinero (pedazo de papel) son, de un lado, ontológicamente subjetivas porque el dinero es un hecho socialmente construido, y es por eso aceptado por todos como medio de pago; de otro lado, son también cognitivamente objetivas, porque no se necesita ningún punto de vista personal para reconocer un billete de diez dólares. Los sentimientos sobre el dinero —«adoro el dinero»— son ontológicamente subjetivos y cognitivamente subjetivos.

Las categorías que se usan en la física y biología solo corresponden a la primera celda. Esta celda también se usa en la economía; sin embargo, en la economía se utiliza también la celda que se encuentra debajo de la primera. Pueden hacerse afirmaciones positivas sobre el dinero, como «la cantidad de dinero ha aumentado en la economía». Luego, las dos celdas de la primera columna son consideradas medibles y falsables en la ciencia económica. La ciencia económica, a diferencia de la física y la biología, utiliza realidades socialmente construidas. Esta característica aumenta la complejidad de la ciencia económica en comparación a las otras disciplinas.

El dinero es ciertamente un ejemplo claro de una variable socialmente construida —ontológicamente subjetiva—. La prueba es que no todo dinero es aceptado en todos los países del mundo. No todas las personas del mundo podrían reconocer un billete de diez soles; sin embargo, todos pueden reconocer que el tamaño de un papel es mayor o menor que el de otro.

La etnicidad es otro ejemplo de un hecho socialmente construido. Tomemos el caso de la raza. La raza como color de la piel es ontológicamente y cognitivamente objetiva; es decir, pertenece a la primera celda (al igual que un pedazo de papel). El color es una característica física de los objetos. Pero raza es también una construcción social, pues es un marcador social de la etnicidad, dado que el color de la piel tiene un significado social. La raza es entonces ontológicamente subjetiva y pertenece a la celda que se encuentra debajo de la primera celda (al igual que el dinero). Por consiguiente, la etnicidad —raza, lengua, religión, cultura, procedencia— también es una categoría socialmente creada. La etnicidad es por consiguiente ontológicamente subjetiva pero cognitivamente objetiva porque pueden hacerse afirmaciones positivas, como «los salarios que reciben los trabajadores negros son, en promedio, menores a los que reciben los trabajadores blancos». El lenguaje, otro marcador social de etnicidad, también pertenece a esta categoría.

Como vimos, el dinero es físicamente un pedazo de papel, pero también es un objeto socialmente construido. El dinero como papel es ontológicamente subjetivo pero cognitivamente objetivo. De manera similar, se puede decir que la raza es como el dinero, mientras que el color físico de la piel es como el papel. Los hechos físicos como las células, colores de la piel y fenotipos —los hechos biológicos— subyacen en el concepto de raza —un hecho socialmente construido—. Ciertamente, los libros de la biología no estudian categorías como la etnicidad. En una sociedad dada, todos saben lo que es el valor de un billete de dinero; similarmente, todos saben cuál es la etnicidad de una persona; es decir todos saben quién es quién en la sociedad. Si este no fuera el caso, la etnicidad apenas existiría como un problema social.

El progreso de las ciencias naturales está basado principalmente en la naturaleza de los objetos que debe medir; se debe también a las innovaciones en los instrumentos de medición. Los telescopios, microscopios y espectrocopios han pasado por un progreso continuo de sofisticación. En efecto, el físico Galison (1997) ha argumentado que los cambios en los paradigmas en la física se han debido principalmente a las innovaciones en los instrumentos de medición. Un nuevo instrumento lleva a nuevas observaciones que pueden falsar un paradigma teórico y producir otro (como se mostró anteriormente). Esta es una hipótesis diferente a la propuesta por Kuhn (1970), quien dijo que son las nuevas ideas políticas y sociales los factores responsables en los cambios de los paradigmas. Ciertamente, la hipótesis de Galison también se aplica a la biología, mientras que la de Kuhn parece aplicarse a la economía.

El progreso de la medición es más difícil de lograr en la economía. Esto por dos razones. Primero, las variables empíricas son mucho más complejas de construir en la economía que en las ciencias naturales, debido a que una parte importante de las variables —tanto endógenas como exógenas— constituyen categorías socialmente construidas. El dinero, el ingreso, el producto total, la pobreza, la desigualdad, la etnicidad, el gobierno, el poder del mercado y la democracia son variables importantes en economía y todas ellas son socialmente construidas.

Segundo, los instrumentos de medición no son tan desarrollados como en las ciencias naturales. En la mayoría de los países, los datos sobre la producción y distribución se construyen aplicando encuestas a las empresas y familias, así como usando las cuentas fiscales y monetarias de los gobiernos. Dado que familias y firmas actúan guiados por la motivación del interés propio, esta motivación estará presente al momento de ofrecer la información. Sobre todo al participar en la producción de un bien público, como la información, los incentivos de las familias y las empresas no son proporcionar la verdadera información, sino la información que «económica y políticamente parecen correctas». Los gobiernos, que también actúan guiados por una motivación del propio interés —la maximización de votos— tampoco tienen los incentivos

para proporcionar y producir la información verdadera, sino la información «políticamente correcta».

Cuando los hechos son socialmente construidos es difícil de medirlos pidiéndoles a las personas que declaren los «hechos». En las sociedades sigma, sociedades que son multiculturales y jerárquicas, el problema es más agudo, ya que cada cultura tiene su realidad socialmente construida, lo cual hace difícil la uniformidad de las variables agregadas. Este problema va más allá de la intención de las personas para decir la verdad o no.

No se ha inventado ningún instrumento tipo telescopio o microscopio para observar la conducta humana en la ciencia económica. Contrariamente a las ciencias naturales, en la ciencia económica, los paradigmas no han sido desafiados por las innovaciones en los instrumentos de medición. Los problemas de medición explican, por lo menos en parte, por qué las teorías económicas tienden a ser inmortales. Si la economía no parece tan científica como las otras ciencias naturales se debe a las diferencias en los instrumentos de medición.

Hacia una ciencia unificada

Considerando la población humana y la Tierra como su envase, ¿existe consistencia entre las teorías de la economía, la física y la biología? Si la respuesta a esta pregunta fuera positiva, las tres disciplinas podrían integrarse en una sola y tendríamos un conocimiento unificado. La tarea específica aquí será analizar si las proposiciones alfa universales de la ciencia económica —aplicables a toda la sociedad humana— son consistentes o no con las teorías de la física y la biología.

El postulado de frontera de producción. Este postulado afirma que los bienes no pueden producirse de la nada; es decir, se necesitan insumos materiales y trabajadores para producirlos. Para cantidades dadas de insumos, el sistema productivo producirá una cantidad determinada de *bienes*, período tras período.

Las leyes de la física están implícitas en este postulado. La gran excepción es la segunda ley de termodinámica, según la cual la cantidad de desorden en el universo aumenta con el tiempo. En la física, el proceso de la producción puede verse como la transformación de baja entropía —materias primas— en alta entropía —residuos y polución—. En este proceso se degrada la energía disponible. Por consiguiente, cuando la producción se repite periodo tras periodo, aunque fuese a la misma escala, la capacidad del sistema de producción también se degrada. En el proceso de la producción, los minerales y otras fuentes de energía se agotan; también se agotan los recursos de la tierra agrícola a través del cultivo continuo, aunque la energía solar es libre y no se agota. En suma, a la larga, este postulado económico contradice la ley de la creciente entropía en el tiempo o Ley de la Entropia. Puede ser válido, como aproximación, solo en periodos cortos.

El postulado de escasez. Este postulado afirma que el deseo humano por los bienes es ilimitado, pero los recursos para producirlos son limitados. La proposición sobre el deseo humano de bienes es consistente con las propensiones humanas postuladas en la biología. «Todo organismo viviente, desde las bacterias hasta las plantas y animales, están dedicados a la búsqueda de algo. En los humanos, la propensión a la búsqueda está amplificada» (Pasternak 2003). Las plantas buscan la luz del sol y tienen la tendencia a crecer hacia él porque el sol es su única fuente de energía; los microbios nadan hacia una fuente de comida; los científicos investigan buscando el conocimiento.

La propensión humana para la búsqueda no es solo «amplificada», sino es ilimitada; es decir, los humanos pueden buscar comida y agua para sobrevivir, así como pueden buscar arte y conocimiento científico para desarrollarse intelectualmente. El psicólogo Abraham Maslow (1970) propuso que las necesidades humanas están jerárquicamente ordenadas, lo que es consistente con la teoría de la búsqueda de la biología. Ambas teorías pueden ser transformadas en una teoría económica del ordenamiento lexicográfico de las preferencias individuales para los bienes. Esto es lo que se ha hecho en este libro.

La economía estándar usualmente supone que los factores de producción primarios son solo el capital físico y el capital humano. Si la economía tuviera más de ambos, podría producir mayores cantidades de bienes, con lo cual implícitamente supone que los recursos naturales son factores redundantes.

De acuerdo con la física, la relación es justamente al revés: los recursos naturales constituyen los factores primarios de producción. Todo puede ser producido, excepto los recursos naturales. La dotación de energía y materia de la humanidad están dadas. La escasez de largo plazo está dada por el tamaño de la Tierra. Pero los recursos naturales son degradados en el proceso de producción —debido a La Ley de la Entropía—. Por ejemplo, el carbón puede ser quemado una sola vez; el petróleo puede ser usado una sola vez. De otro lado, dado el tamaño de la Tierra, la Ley de la Entropía es la causa última de la escasez en la producción de bienes.

El mismo proceso de la producción es visto de distinta manera en cada ciencia. En el proceso económico se produce no solo *bienes*, sino también otros «productos» que igualmente salen del proceso, como deshechos, basura y contaminantes del medio ambiente. La teoría económica estándar hace abstracción de estos últimos. La física, en cambio, considera como producto todo lo anterior. La Ley de la Entropía o, más propiamente, la teoría de la entropía predice, por lo tanto, que el ambiente físico será degradado en el proceso de producción de bienes. Esto es consistente con lo que nosotros observamos. Después de un período largo de crecimiento del producto mundial, el planeta Tierra está experimentando ya un problema de contaminación. La Ley de la Entropía es ignorada en la economía estándar, pero la realidad ha mostrado que esta ley es demasiado importante como para ser ignorada.

En el largo plazo, existe un dilema fundamental entre la producción presente y la producción futura. Las preferencias de la población actual son por más bienes ahora que para las generaciones futuras. Esta motivación queda revelada por la obsesión con que firmas, hogares, gobiernos, organizaciones multilaterales actúan y demandan

maximizar el crecimiento del producto actual. No existe ningún mecanismo de precios que pueda resolver este dilema, este problema económico fundamental del la especie humana, pues las preferencias de las generaciones futuras no cuentan. Y, lamentablemente, la madre naturaleza no tiene cajero. Debido a esta inconsistencia entre las teorías de la economía y las ciencias naturales, no existe unidad del conocimiento —todavía— entre las ciencias.

El postulado de incertidumbre. Este postulado afirma que los humanos enfrentan percances aleatorios a las dotaciones de sus recursos. Parte de estos percances vienen de la sociedad misma y parte proviene de la madre naturaleza, como inundaciones, sequías, terremotos y epidemias. Estas variables exógenas son el resultado de sistemas caóticos que existen en el mundo físico o en el mundo biológico.

Este postulado es consistente con lo que sostienen las teorías modernas de la física. Con el desarrollo de la teoría cuántica, la física es ahora una disciplina más compleja que antes, pues las relaciones en el mundo subatómico son imprevisibles o inciertas. El ganador del Premio Nobel, Ilya Prigogine (1997) ha escrito un libro con el sugestivo título de *The End of Certainty* (que podría traducirse como (El fin de la certeza en la física) en el cual presenta la nueva complejidad de la física. La física está volviéndose una ciencia más compleja, más cercana a la economía.

Los otros postulados de la ciencia económica —institucionalidad, racionalidad y las condiciones iniciales— no contradicen teorías de la física o de la biología. Entonces, la economía, la biología y la física podrían ser consideradas ciencias complementarias, que podrían generar una teoría unificada, si solo la Ley de la Entropía fuera integrada en la economía. Algunos progresos han sido hechos en esta dirección, como el realizado por el economista Georgescu-Roegen (1971), considerado como uno de los pioneros en este campo.

Las comparaciones hechas entre la economía, la física y la biología pueden resumirse como sigue. Sobre la naturaleza del proceso que estudian, la economía se parece más a la biología que a la física, pues los procesos económicos y biológicos asumen la existencia de variables

exógenas. Sobre el universalismo ontológico, solo la física puede suponer un mundo único y una teoría única; la economía y la biología son ciencias con teorías parciales y unificadas. Sobre la metodología, las tres ciencias pueden utilizar la epistemología de la falsación popperiana para refutar las teorías; sin embargo, la física no puede utilizar el método alfa-beta, pues en el proceso físico no existen variables exógenas. Sobre las mediciones, la economía es diferente de la física y la biología. Ningún instrumento como el microscopio, telescopio o espectroscopio existe en la economía. Las mediciones en la economía son relativamente imperfectas.

Finalmente, sobre la unidad de conocimiento entre estas ciencias, las teorías en la economía, en la física y en la biología son fundamentalmente consistentes entre ellas. Existe una excepción importante, sin embargo. La ciencia económica ignora el efecto de la Ley de la Entropía en el proceso económico. La teoría económica del crecimiento supone que el crecimiento del producto puede ser perpetuo, pero La Ley de la Entropía niega esta posibilidad.

Glosario

Bien público: bien que se produce y consume colectivamente, como por ejemplo, las carreteras.

Bien inferior: bien cuyo consumo disminuye cuando el ingreso real de los consumidores aumenta.

Bien normal: bien cuyo consumo aumenta cuando el ingreso real de los consumidores aumenta.

Capital físico: stock de herramientas y maquinaria que se utilizan en el proceso productivo.

Capital fijo: igual que el capital físico.

Capital circulante o capital de trabajo: stock de dinero —propio o crédito— que las firmas necesitan para llevar a cabo el proceso productivo en el corto plazo.

Capital humano: stock de conocimientos productivos que se encuentran incorporados en los trabajadores.

Causalidad: relación entre variables endógenas y exógenas de una teoría; es una proposición beta.

Coeficiente de Gini: medida del grado de desigualdad de una distribución; varía entre 0 y 1, desde la completa igualdad (0) hasta la completa desigualdad (1), cuando solo una persona concentra todo el ingreso o todo el recurso.

Competencia perfecta: estructura de un mercado que se compone de muchos compradores y vendedores, de modo que nadie tiene poder para fijar el precio; y, todos los participantes son precio-aceptantes.

Competencia imperfecta: estructura de un mercado donde no todos los participantes son precio-aceptantes; es decir, algunos tienen poder de fijar el precio.

Costo promedio: costo por unidad de producto producido.

Costo marginal: incremento en el costo total de la firma debido al aumento en la producción en una unidad del bien producido. En un mercado de competencia perfecta se le supone creciente —es el otro lado de los rendimientos decrecientes— y da lugar a la curva de oferta de la firma.

Curva de demanda: relación inversa entre el precio de un bien y las cantidades que los agentes desean comprar en el mercado (para valores dados de las variables exógenas). Cambios en la demanda significan desplazamientos de la curva hacia adentro o hacia fuera.

Curva de oferta: relación directa entre el precio de un bien y las cantidades que los agentes desean vender en el mercado (para valores dados de las variables exógenas). Cambios en la oferta significan desplazamientos de la curva hacia adentro o hacia fuera.

Curva de productividad promedio del trabajo: relación inversa entre el producto por trabajador y la cantidad de trabajadores, dado los valores del stock de capital y la tecnología.

Curva de productividad marginal del trabajo: relación inversa entre la productividad marginal del trabajo y la cantidad de trabajadores, dado los valores del stock de capital y la tecnología.

Desigualdad inicial: desigualdad en la distribución de las dotaciones individuales de activos económicos y sociales.

Desarrollo económico: combinación del nivel de ingreso medio y el grado de igualdad en la distribución del ingreso, la cual representa el nivel de progreso social de una sociedad.

Equilibrio: solución del intercambio entre los agentes económicos, donde ningún agente tiene el poder y el deseo de cambiar la solución.

Equilibrio microeconómico: solución de precios o cantidades que los agentes eligen individualmente, dados los valores de las variables exógenas que enfrentan.

Equilibrio parcial: solución de precio y cantidad en un mercado determinado que los agentes tiran interactuando entre ellos.

Equilibrio general: solución de precios y cantidades en todos los mercados que los agentes determinan interactuando entre ellos.

Equilibrio estático: Los valores de las variables endógenas se repiten periodo tras periodo si los valores de las variables exógenas se mantienen fijos.

Equilibrio dinámico: Los valores de las variables endógenas cambian a través del tiempo en una trayectoria determinada, manteniendo fijos los valores de las variables exógenas.

Estructura de mercado: composición de compradores y vendedores en el intercambio con distintos grados de poder para fijar el precio; cuando nadie puede fijar el precio, la estructura se denomina competencia perfecta; cuando un solo vendedor puede fijar el precio, se denomina monopolio; cuando varios vendedores pueden fijar el precio, se denomina oligopolio.

Exceso de demanda: al precio que rige en el mercado de un bien, los agentes no pueden comprar las cantidades que desean.

Exceso de oferta: al precio que rige en el mercado de un bien, los agentes no pueden vender las cantidades que desean.

Falacia de composición: principio lógico según el cual una relación, que es válida para un individuo, no lo es para el conjunto de individuos.

Función de producción: relación que se da en el proceso productivo y que indica que la cantidad producida depende de la cantidad de trabajo y de la cantidad de capital que se utiliza.

Ley de Warlas: si en un sistema compuesto de n mercados walrasianos, n-1 mercados están en equilibrio, entonces el enésimo mercado estará necesariamente en equilibrio.

Macroeconomía: teoría que busca explicar la producción y distribución de una sociedad. Teoría de equilibrio general.

Microeconomía: teoría que busca explicar el comportamiento de un actor social, considerado individualmente.

Mercado: lugar o espacio en el cual compradores y vendedores llevan a cabo el intercambio de bienes bajo normas sociales particulares. El lugar no es necesariamente físico.

Mercado walrasiano: el precio nominal es flexible, es decir, puede subir o bajar hasta que se llegue a un precio al cual los agentes puedan comprar o

vender todas las cantidades que deseen intercambiar. Este es el precio de equilibrio. Y a este precio nadie se queda sin intercambiar las cantidades que desea; el mercado se «limpia».

Mercado no walrasiano: el precio nominal es totalmente fijo, o puede ser parcialmente fijo, cuando existe algún tipo de rigidez para poder bajar o subir debido a ciertas normas sociales. También el concepto se aplica cuando se trata del precio relativo. El equilibrio en el mercado se da, entonces, con exceso de demanda o de oferta.

Modelo de una teoría: conjunto de proposiciones alfa de una teoría y un conjunto de proposiciones auxiliares que sean consistentes entre si y que permitan derivar proposiciones beta. Una teoría es una familia de modelos.

Ontología universalista: supuesto que establece que la realidad es única y, por lo tanto, universal, en lugar de suponerla como compuesta de varios tipos.

Plazos: tiempo lógico —y no cronológico— para el análisis. El análisis del proceso económico considerando que el stock de capital fijo se llama «corto plazo» y cuando el stock de capital es variable se llama «largo plazo».

Proceso económico: actividad humana de producir y distribuir bienes que tienen una duración determinada, elementos endógenos y exógenos, y que se repite periodo tras periodo

Precio nominal: la cantidad de dinero que se tiene que dar en el mercado a cambio de una unidad del bien, por ejemplo, soles por kilo de papa.

Precio relativo: la cantidad de un bien A —que no sea dinero— que se tiene que dar en el mercado a cambio de una unidad de otro bien B, por ejemplo, litros de leche por kilos de papa.

Precio walrasiano: precio de equilibrio en un mercado walrasiano.

Precio no walrasiano: precio que rige en un mercado no walrasiano.

Principio de la endogenización creciente: a medida que el nivel de análisis pasa del equilibrio micro, al parcial y al general, las variables exógenas se van transformando en variables endógenas.

Principio de la falsabilidad: una teoría es científica cuando en principio puede ser falsa, es decir, cuando en principio puede ser refutada por los datos de la realidad.

Principio de la razón insuficiente: si una hipótesis empírica no es rechazada estadísticamente, no hay razón alguna para rechazarla y puede ser aceptada como válida.

Principio de Moore: conjunto de teorías parciales puede dar lugar a una teoría unificada si los supuestos de las teorías parciales dan lugar a un núcleo de supuestos que son consistentes entre sí.

Prima: premio que los compradores o vendedores pagan por encima del precio de mercado de un bien.

Problema de Gödel: un sistema no puede comprenderse (explicarse o modificarse) a sí mismo. Por ejemplo: el diccionario no puede definir todas sus palabras; el cerebro humano no puede estudiarse a sí mismo.

Problema de Olson: si los individuos actúan guiados por su propio interés, no podrán producir un bien público, pues el interés de todos se convierte en el interés de nadie.

Problema de la falacia de composición: una proposición que es cierta para un individuo, no es cierta para el agregado de individuos.

Productividad promedio del trabajo: cantidad de producto en una unidad de tiempo dividido por la cantidad de trabajadores utilizada en el proceso productivo; es decir, es el producto por trabajador.

Productividad marginal del trabajo: cantidad de producto que aumenta en una unidad de tiempo debido al aumento de un trabajador en el proceso productivo, manteniendo constante el stock de capital y la tecnología.

Proposición alfa: supuesto primario de una teoría; no es observable.

Proposición beta: predicción empírica que se deriva lógicamente del conjunto de proposiciones alfa de una teoría, o de un modelo de la teoría, y que es observable y, por lo tanto, empíricamente refutable.

Rendimientos a escala constantes: al aumentar cantidad de factores utilizados en el proceso productivo en una proporción dada, la cantidad del producto también aumenta en esa proporción; es decir, el doble de factores produce el doble del producto.

Rendimientos decrecientes del trabajo: relación inversa entre productividad laboral y cantidad de trabajadores, para cantidades dadas del stock de capital y tierra; es decir, el doble de trabajadores no produce el doble de producto, sino menos.

Rendimientos decrecientes ricardianos: relación inversa entre productividad laboral y cantidades de trabajadores, para cantidades dadas del stock de capital y tierra, y para calidades cada vez menores de tierras y otros recursos naturales.

Salario nominal: cantidad de dinero que se intercambia en el mercado laboral por una unidad de servicio laboral (por ejemplo, soles por semana de cuarenta horas de trabajo).

Salario real: poder de compra del salario nominal, por ejemplo, kilos de papa por semana de cuarenta horas de trabajo.

Salario real de eficiencia: salario que crea incentivos en los trabajadores para que incurran en la indisciplina laboral y en el poco esfuerzo en la firma.

Superpoblación o sobrepoblación: cantidad de capital por trabajador en una sociedad que es baja, tal que la productividad marginal del total de trabajadores es cero, cerca a cero o por debajo del salario de subsistencia. Si es cero significa que la disminución de un trabajador, o de un grupo de trabajadores, no modificaría la cantidad de producto que produce la economía.

Teoría: conjunto de supuestos primarios —no observables— sobre la realidad que lleva a construir una sociedad abstracta y más simple que la sociedad real, y que permitirá comprender la realidad en sus aspectos esenciales. Es un conjunto de proposiciones alfa.

Trabajadores Z: aquellos que tienen la dotación más baja en activos políticos, así como también la dotación más baja en capital humano.

Variable endógena: variable observable cuyo valor es resultado del proceso económico.

Variable exógena: variable observable cuyo valor se determina fuera del proceso económico y cuyos cambios afectan los valores de equilibrio de las variables endógenas.

Variable nominal: variables expresada en unidades del dinero, como la cantidad de dinero de la economía.

Variable real: variable expresada en unidades físicas —como producto y empleo— o variables nominales expresadas en unidades de otro bien —como salario real—.

Bibliografía

ALESINA, Alberto y Roberto PEROTTI
1996 «Income Distribution, Political Instability, and Investment». En *European Economic Review,* 40, pp. 1203-1228.

BARRO, Robert
1997 *Macroeconomics.* Quinta edición. Cambridge: MIT Press.

BARRO, Robert y Xavier SALA-I-MARTIN
1995 *Economic Growth.* Nueva York: McGraw Hill.

BLANCHARD, Oliver
2004 *Macroeconomics.* Cuarta edición. Cambridge: MIT Press.

BOURGUIGNON, Francois
2000 «Crime, Violence and Inequitable Development». En *Annual World Bank Conference on Development Economics 1999.* Washington D.C.: The World Bank.

DARITY, William y Jessica NEMBHARD
2000 «Racial and Ethnic Economic Inequality: The International Record». En *American Economic Review,* 90 (2), mayo, pp. 308-311.

DEININGER, Klaus y Lyn SQUIRE
1996 «A New Data Set Measuring Inequality». En *The World Bank Economic Review*, vol. 10, N° 3, setiembre.

FAJNZYLBER, Pablo, Daniel LEDERMAN y Norman LOAYZA
1999 «What Causes Violent Crime». En *European Economic Review,* 46, pp. 1323-1358.

Figueroa, Adolfo
2003 *La sociedad sigma. Una teoría del desarrollo económico.* Lima y México D.F.: Fondo Editorial de la Pontificia Universidad Católica del Perú y Fondo de Cultura Económica.
2006 «El problema del empleo en una sociedad sigma». Documento de Trabajo N° 249, Departamento de Economía, Pontificia Universidad Católica del Perú, setiembre.

Galison, Peter
1997 *Image and Logic. A Material Culture of Microphysics.* Chicago: The University of Chicago Press.

Garraty, John
1978 *Unemployment in History.* Nueva York: Harper Colophon Books.

Georgescu-Roegen, Nicholas
1971 *The Entropy Law and the Economic Process.* Cambridge: Harvard University Press.

Hall, Gillette y Anthony Patrinos
2005 *Indigenous Peoples, Poverty, and Human Development in Latin America.* Londres: Palgrave.

Hawking, Stephen
1996 *A Brief History of Time.* Nueva York: Bantam Books. Traducido como *Historia del tiempo* (Madrid: Alianza 2005).

Jones, Charles
1998 *Introduction to Economic Growth.* Londres: W.W. Norton. Traducido como *Introducción al crecimiento económico* (México D.F.: Pearson Education).

Keynes, John
1936 *The General Theory of Employment, Interest and Money.* Londres: Macmillan. Traducido como *Teoría general de la ocupación, el dinero y el interés* (Mexico D.F.: Fondo de Cultura Económica 1971).

Kuhn, Tomas
1970 *The Structure of Scientific Revolutions.* Chicago: The University of Chicago Press.

Li, Hongyi, Lyn Squire, Heng-fu Zou
1998 «Explaining International and Intertemporal Variations in Income Inequality», En *Economic Journal,* 108(1), enero, pp. 26-43.

Lucas, Robert
1990 «Why Doesn't Capital Flow from Rich to Poor Countries». En *American Economic Review* 80 (2), pp. 92-96.

Maddison, Angus
1995 *Monitoring the World Economy, 1820-1992.* París: Development Centre, OECD.

Markusen, James
2002 *Multinational Firms and the Theory of International Trade.* Cambridge: MIT Press.

Marx, Karl
1867 *Das Kapital.* Traducido al castellano como *El capital* (Buenos Aires: Siglo XXI, 1975).

Maslow, Abraham
1970 *Motivation and Personality.* Segunda edición. Nueva York: Harper & Row Publishers.

Mayr, Ernest
1991 *One Long Argument. Charles Darwin and the Genesis of Modern Evolutionary Thought.* Cambridge: Harvard University Press.

Milanovic, Branco
2006 *La era de las desigualdades. Dimensiones de la desigualdad internacional y global.* Madrid: Editorial Sistema.

Muller, Edward
1997 «Economic Determinants of Democracy». En: *Inequality, Democracy, and Economic Development,* M. Midlarsky (editor). Cambridge: Cambridge University Press.

Nayyar, Deepak
2003 «The Political Economy of Exclusion and Inclusion: Democracy, Markets, and People». En: *Development Economics and Structural Macroeconomics. Essays in Honor of Lance Taylor,* A. Dutt y J. Ros (editores). Cheltenham: Edward Elgar Publishers.

Olson, Mancur
1965 *The Logic of Collective Action. Public Goods and the Theory of Groups.* Cambridge: Harvard University Press.

Popper, Karl
1959 *The Logic of Scientific Discovery.* Londres: Hutchinson.
1993 «Evolutionary Epistemology». En: *Contemporary Readings in Epistemology,* M. Goodman y R. Snyder (editores). Engelwood Cliffs: Prentice Hall.

Prigogine, Ilya
1997 *The End of Certainty. Time, Chaos, and the New Laws of Nature.* Nueva York: The Free Press.

Psacharopoulos, George y Harry Patrinos
1994 *Indigenous People and Poverty in Latin America. An Empirical Analysis.* Washington D.C.: World Bank.

Rawls, John
1971 *A Theory of Justice.* Cambridge: Harvard University Press.

Ricardo, David
1821 *Principles of Political Economy and Taxation.* Traducido como *Principios de economía política y tributación* (México D.F., Fondo de Cultura Económica 1973).

Sen, Amartya
1992 *Inequality Reexamined.* Cambridge: Harvard University Press.

Silva, Nelson
2001 «Race, Poverty, and Social Exclusion in Brazil». En: *Social Exclusion and Poverty Reduction in Latin America,* E. Gacitúa, C. Sojo y S. Davies (editores). Washington D.C.: The World Bank.

Smith, Adam
1776 *An Enquiry into the Nature and Causes of the Wealth of Nations.* Traducido como (Barcelona: Orbis 1983).

Stewart, Frances
2001 *Horizontal Inequalities: A Neglected Dimension of Development.* Helsinki: United Nations University.

Walras, Leon
1954 *Elements of Pure Economics.* Nueva York: A. Kelly Publishers. Originalmente publicado en francés en 1883.

World Bank
2001 *World Development Report 2000/2001. Attacking Poverty.* Oxford: Oxford University Press.

Índice analítico

Impreso en los talleres de
LITHO & ARTE SAC
Jr. Iquique 046 - Breña
Teléfonos: 332-1989 / 332-8397 / 332-9077
E-mail: lithoarte@speedy.com.pe
Marzo del 2008