



UNIVERSIDAD  
DE SANTIAGO  
DE CHILE

Artículo de Investigación  
<https://doi.org/10.35588/cc.v3i2.5816>



**Lucas Miranda Baños**

lucas.mirba@gmail.com

<https://orcid.org/0000-0003-1242-0910>

CIECE-FCE-UBA, Argentina.

Artículo recibido: 01 de noviembre de 2022  
Artículo aceptado: 29 de diciembre de 2022  
Artículo publicado: 31 de diciembre de 2022



[CC BY, Lucas Miranda Baños, 2022]

## Modelos estadísticos y explicación del error inferencial

*Statistical Models and Explanation of Inferential Error*

### Resumen

La tesis principal de este trabajo es que los modelos estadísticos facilitan explicaciones acerca de los errores que pueden cometer nuestros procedimientos de inferencia de datos a fenómenos, en el contexto de la estadística clásica o frecuentista. La explicación del error es una operación que, a su vez, ayuda a evitarlo en donde es evitable e intolerable; tolerarlo donde es inevitable y tolerable; y suspender el juicio donde es inevitable e intolerable. Todas estas operaciones son elementos de una práctica generadora de evidencia confiable. Se ilustra esta tesis con el modelo lineal en econometría, el estimador de mínimos cuadrados ordinarios, el estimador y modelo de variables instrumentales y con los conceptos relacionados con las pruebas de hipótesis. Se pone especial énfasis en las propiedades de insesgadez, consistencia y precisión como caracterizaciones del tipo de error que se puede cometer usando un estimador. Se sostiene que los modelos explican identificando patrones de dependencia contrafáctica entre características del mecanismo estocástico representadas por el modelo y características de los procedimientos inferenciales.

**Palabras clave:** Modelos estadísticos, Error, Procedimientos inferenciales, Explicación causal, Patrones de dependencia contrafáctica.

### Abstract

The main thesis of this work is that statistical models provide explanations about the errors that our data inference procedures from data to phenomena can make, in the context of classical or frequentist statistics. The explanation of the error is an operation which in turn, helps to avoid it where it is avoidable and intolerable; tolerate it where it is unavoidable and tolerable; and suspend judgment where it is unavoidable and intolerable. All these operations are elements of a reliable evidence-generating practice. This thesis is illustrated with the linear model in econometrics, the ordinary least squares estimator, the estimator and model of instrumental variables, and with the concepts related to hypothesis testing. Special emphasis is placed on the properties of unbiasedness, consistency, and precision as characterizations of the type of error that can be made using an estimator. It is argued that models explain by identifying patterns of counterfactual dependence between features of the stochastic mechanism represented by the model and features of the inferential procedures.

**Keywords:** Statistical models, Error, Inferential procedures, Causal explanation, Counterfactual dependency patterns.

## 1. Introducción<sup>1</sup>

La tesis principal de este trabajo es que los modelos estadísticos —en la estadística frecuentista o estadística del error (Mayo y Spanos, 2011)— facilitan explicaciones acerca de los errores que pueden cometer nuestros procedimientos de inferencia de datos a fenómenos. Un modelo estadístico ayuda a responder por qué y cómo pueden errar estas inferencias, permitiendo explicar el error. La explicación del error es una operación que a su vez ayuda a evitarlo<sup>2</sup> en donde es evitable e intolerable; tolerarlo donde es inevitable y tolerable; y suspender el juicio donde es inevitable e intolerable<sup>3</sup>. Todas estas operaciones son elementos de una práctica generadora de evidencia confiable.

Este planteamiento no aspira a ser demasiado original. Tiene un aspecto inconfundiblemente popperiano y ha sido desarrollado por Deborah Mayo. De acuerdo con Mayo, “una manera de describir cómo los métodos estadísticos funcionan, llegué a ver, es que nos permiten, literalmente, aprender del error” (Mayo, 1996, p.xii). El objetivo de este artículo es enfatizar que aprendemos de los errores porque podemos explicarlos, y los modelos estadísticos ayudan a ello. Se defenderá que los modelos facilitan explicaciones causales del error en el sentido de Woodward (2003), mediante la identificación de patrones de dependencia contrafáctica entre características del mecanismo generador de datos (representadas por un modelo) y características del error de los procedimientos inferenciales. La explicación del error a través de sus causas facilita su gestión, evitándolo cuando es evitable y reconociéndolo cuando es inevitable.

El trabajo puede concebirse respondiendo tres preguntas: la primera es ¿qué hacen los modelos estadísticos? La respuesta que se dará es que, entre otras funciones, una fundamental es proveer explicaciones acerca de las posibles formas de error inferencial. En la tercera sección se describirá la manera en que la inferencia estadística en la tradición frecuentista se basa en modelos, siguiendo los planteamientos de Spanos (2006a, 2021). En la cuarta sección se ilustrará con ejemplos de la estadística cómo se explican distintas formas de error a la hora de realizar inferencias. La segunda pregunta es ¿para qué explican el error? La respuesta será que lo hacen para evitarlo y tolerarlo, lo cual es fundamental para producir buena evidencia. La segunda sección trazará históricamente la idea de que la buena evidencia es aquella que se pone en guardia ante las fuentes de error. La cuarta sección ilustrará con el sesgo de omisión de variables y el uso de variables instrumentales el rol que cumplen los modelos estadísticos para evitar los errores que surgen de la endogeneidad. La tercera

---

<sup>1</sup> Agradezco a los revisores por los comentarios que contribuyeron a mejorar este artículo. Todas las traducciones del inglés de los pasajes citados son mías.

<sup>2</sup> Toda forma de error no se puede evitar, pero ciertas formas de error (como el sesgo y la inconsistencia) sí se pueden evitar y algunas formas de error, como la imprecisión, se pueden aminorar. Cuando en el trabajo se hable de evitar el error, se debe tener en cuenta este *caveat* y no pensar que se considera posible evitar toda forma de error.

<sup>3</sup> Esta relación instrumental se comprende mejor si se distingue entre la explicación científica a nivel factual, donde los *explananda* son los fenómenos sustantivos que estudia la ciencia y la explicación científica en un meta-nivel metodológico, donde los *explananda* son aspectos de la actividad científica. La explicación en el meta-nivel metodológico tiene un rol instrumental para la explicación en el nivel factual. Agradezco a Nélica Gentile por esta observación.

pregunta es ¿cómo lo hacen? en la quinta sección se responderá que explican identificando patrones de dependencia contrafáctica en el sentido de Woodward (2003).

## 2. Error y evidencia

Desde los comienzos de la filosofía occidental, con Parménides, la reflexión acerca del conocimiento ha venido acompañada por una teoría del error que hace falta evitar para obtener conocimiento. La diosa le enseña al eleata dos vías: la de la persuasiva verdad y “aquella por la cual los mortales que nada saben yerran” con el objetivo de que el “hábito inveterado” no lo lleve a conducir por esa vía el “ojo sin meta” (Gallop, 1984). No es una exageración decir que la metáfora de la línea y la alegoría de la caverna en La República de Platón proveen una metafísica del error (Platón, 2004). En la modernidad, Bacon presenta en el *Novum Organum* una teoría de los ‘ídolos’ o nociones falsas que ponen obstáculos a la ciencia y llama a ponerse en guardia contra ellos (Bacon, 2019). La duda metódica de las Meditaciones Metafísicas de Descartes recorre diversas fuentes de error como un medio para alcanzar el principio indubitable, el *cogito* (Descartes, 2013).

Un hilo conductor de esta tradición epistemológica, que Popper ha denominado la doctrina de la verdad manifiesta (Popper, 2014), es que se concibe el error como algo capaz de ser evitado completamente o, inversamente, que la certeza es un ideal alcanzable. En el siglo XIX este ideal se desvanece y es reemplazado por una visión falibilista de la ciencia, en donde el error ya no puede ser completamente eliminado (Laudan, 1984, p.83). El conocimiento (y la evidencia) ya no puede ni debe ser infalible. Este giro conlleva la necesidad de profundizar en la teoría del error, distinguiendo distintos tipos de error, determinando cuáles son tolerables y cuáles no, y cómo evitarlos o aminorarlos donde es posible.

No es casualidad que la aplicación de la estadística moderna a la investigación empírica cobre fuerza en el siglo XIX y en el XX. Un trabajo seminal en esta área es aquel en el cual Gauss (1995[1823]) establece la fórmula de la función de densidad de la distribución normal. El título de ese trabajo es el de *Teoría de la combinación de observaciones sujetas al mínimo error*, en el cual aborda el problema de los errores de medición, hace una serie de distinciones (e.g. error cierto y aleatorio, error medio, error probable) y desarrolla el método de estimación de mínimos cuadrados. Se trata de un trabajo que apunta al entendimiento y minimización del error cuando se lo entiende como un fenómeno aleatorio. En trabajos como el de Galton y Pearson, Neymann y Fisher a fines del siglo XIX y principios del siglo XX la estadística se desarrolla y encuentra más aplicaciones científicas, entre las que se encuentra la econometría.

Estas disciplinas introducen una amplia familia de conceptos de error: el término del error  $U$  en el modelo lineal, el error estándar de una variable aleatoria, errores de tipo I y II en las pruebas de hipótesis, sesgos e imprecisión de estimadores, error estándar de un estimador, error de medición de una variable, entre otros. Desde el punto de vista de este trabajo, la existencia de esta variedad de conceptos no es una casualidad. De acuerdo a Mayo y Spanos, lo que caracteriza a la estadística del error (i.e. denominación que le dan a la estadística frecuentista<sup>4</sup> y al enfoque filosófico que la

<sup>4</sup> La estadística frecuentista trata la extracción de la muestra como un experimento aleatorio repetible sobre los cuales se construyen variables aleatorias a las cuales se les asignan probabilidades. El valor de los parámetros poblacionales, a

acompaña) es una particular respuesta a la pregunta “¿cómo obtenemos conocimiento confiable acerca del mundo a pesar de la incertidumbre y amenazas de error?” (2011, p. 153). De esta manera, identificar y explicar las distintas amenazas de error en el razonamiento a partir de datos es una tarea necesaria para la obtención de conocimiento confiable.

La relación entre la evidencia y el error que se reconocerá es la siguiente: la evidencia *prima facie* para una hipótesis entendida de manera contextualista es cualquier cosa que sea relevante saber para creer en la hipótesis (Reiss, 2008, p.7). La evidencia *prima facie* no será buena evidencia a menos que descartemos algunas fuentes importantes de error. Desde un punto de vista falibilista no se podrá descartar todas las fuentes de error y será importante explicitar las diferencias entre el error que se puede descartar y el que no. Por ejemplo, si un síntoma es evidencia *prima facie* de determinada enfermedad porque la enfermedad explica el síntoma, una fuente de error es la existencia de explicaciones alternativas del síntoma. Para que el síntoma se considere buena evidencia debe haberse descartado estas explicaciones alternativas. Otro ejemplo más apropiado a nuestro objetivo es la posibilidad de un sesgo en la estimación de un parámetro. Para que la estimación sea buena evidencia del valor del parámetro debe haberse argumentado contra la presencia de un sesgo en su estimación.

Dado que los errores que podemos cometer al establecer evidencia para una hipótesis difieren en distintos campos, los estándares de buena evidencia variarán de campo en campo. Los errores asociados a la utilización de instrumentos ópticos en astronomía son muy distintos a los errores que se pueden cometer diagnosticando una enfermedad. Por este motivo las teorías epistemológicas modernas —como el racionalismo y el empirismo— que pretenden erigirse sobre la base de una *teoría general* del error tienen menor importancia en un mundo falibilista, en donde la consciencia de las particularidades del error se profundiza.

Otra consecuencia de este planteamiento es que, en distintos campos, es concebible el progreso teórico en la obtención de mejor evidencia. Este progreso estará dado por avances en la teoría del error propia de ese campo. Por ejemplo, si en cierto momento en medicina no se tenía conocimiento de que determinados síntomas eran compatibles con una variedad de diagnósticos, al obtener este conocimiento y el de mecanismos para descartar diagnósticos alternativos se llegará a una situación en donde la calidad de la evidencia mejorará. Análogamente, algunos avances en econometría pueden ser considerados productos de una mayor consciencia de fuentes de error y de los medios para evitarlos.

No obstante, el falibilismo surge del reconocimiento de que el error no puede ser nunca completamente evitado. Se impone con ello la necesidad de distinguir el error tolerable del intolerable y el evitable del inevitable. La metodología y la teoría del error específica que requiere cada campo se impone una tarea análoga a lo que reza la oración de alcohólicos anónimos: *señor, dame serenidad para aceptar los errores que no puedo cambiar, coraje para cambiar los que puedo y sabiduría para conocer la diferencia*. En última instancia, si se reconoce que hay amenazas de

diferencia de la estadística bayesiana, no es considerado una variable aleatoria. Sobre esta base se construyen procedimientos inferenciales como las pruebas de hipótesis, en donde se reconocen distintas probabilidades de error a la hora de aceptar o rechazar una hipótesis.

error ni evitables ni tolerables, la alternativa es la suspensión del juicio, el reconocimiento de que la evidencia es insuficiente. Es por esto que consideraremos las caracterizaciones del error como propias tanto de la buena como de la mala evidencia. Los procedimientos que producen buena evidencia pueden errar, lo que debe caracterizarse es cómo lo hacen.

### 3. Inferencia estadística basada en modelos

En esta sección se seguirá de cerca la caracterización que hace Spanos (2006a, 2012, 2021) de la estadística frecuentista como una basada en modelos. Desde esta visión, el problema central de la estadística inferencial es cómo inferir características de la población a partir de datos muestrales. A partir de Fisher (1922), la estadística frecuentista trata a la muestra  $x_0 = (x_1, \dots, x_n)$  como una realización típica de un mecanismo estocástico  $\{X_t, t \in \mathbb{N}\}$ . Este mecanismo generador de datos desde el que se concibe la muestra es representado mediante un modelo  $M_\theta(x)$  respecto del cual se hacen distintos supuestos (Spanos, 2021, p.399). Ese modelo plantea posibles distribuciones conjuntas de las variables muestrales, un espacio de parámetros posibles y hace supuestos acerca de las variables aleatorias y su distribución conjunta. La inferencia busca, usando la información muestral y los supuestos del modelo, acotar lo más que se pueda el espacio de parámetros posibles, en lo posible a un solo valor (Spanos, 2021, p. 400).

Ejemplos de modelos estadísticos son el modelo normal, el modelo lineal y el modelo de Bernoulli. Estos modelos a su vez pueden ser especificados de distintas maneras. Por ejemplo, un modelo lineal puede contener distintas variables independientes, relaciones cuadráticas o lineales, logaritmos en las variables independientes y/o dependiente. El problema de la especificación de un modelo responde a la pregunta de qué mecanismo estocástico es realizado por determinados datos. La misma estará guiada tanto por consideraciones teóricas como por el examen de los datos<sup>5</sup> (Spanos, 2021, p.400). Las inferencias que se hagan partiendo del modelo dependerán crucialmente de que el modelo sea estadísticamente adecuado. Si los supuestos no se cumplen y el modelo es inadecuado, la inferencia no será confiable.

La teoría estadística permite deducir proposiciones inferenciales a partir de los supuestos del modelo estadístico. Estas proposiciones inferenciales se refieren a propiedades de procedimientos inferenciales como estimadores, predictores, intervalos de confianza y pruebas de hipótesis (Spanos, 2021, p.401). Un ejemplo de estas deducciones es la demostración del teorema de Gauss-Markov y un ejemplo de proposición inferencial es el teorema de Gauss-Markov mismo, de lo cual se hablará más adelante. Lo que se sigue deductivamente, sostendremos, son caracterizaciones del error que estos procedimientos inferenciales pueden cometer. Ejemplos de estas caracterizaciones son la atribución de propiedades como insesgadez, consistencia, eficiencia de los estimadores. Otro ejemplo lo constituyen proposiciones acerca de las distribuciones y probabilidades de error de las

---

<sup>5</sup> Según Spanos la especificación de un modelo para determinados datos es el resultado de un razonamiento abductivo en donde la afirmación de que determinado modelo explica los datos, es la conclusión del razonamiento.

distintas pruebas de hipótesis como son los errores de tipo I y II, el poder y el  $p$ -valor<sup>6</sup> de una prueba de hipótesis.

El uso de datos específicos habilita -en conjunto con las proposiciones acerca de los procedimientos inferenciales- resultados inferenciales específicos. Es importante notar que las caracterizaciones del error pueden ser acerca de los resultados inferenciales específicos derivados de una muestra particular o acerca de propiedades genéricas de los procedimientos inferenciales. Un ejemplo de lo primero es la varianza de un estimador. En la sección siguiente nos referiremos a ella, pero acá cabe señalar que es una medida de la imprecisión de un estimador y su determinación es función de una muestra específica. Por otro lado, hay propiedades genéricas de los procedimientos inferenciales que constituyen caracterizaciones del error. La insesgadez de un estimador es un ejemplo de ello. Las probabilidades de error de tipo I y II son probabilidades de las reglas de pruebas de hipótesis, no del resultado de una prueba particular. Estos son ejemplos de características genéricas del error que pueden cometer los procedimientos inferenciales.

Hasta aquí la caracterización de la estadística frecuentista que hace Spanos. Lo que se agregará es la tesis de que entre los supuestos que especifican un modelo y las formas de errar de los procedimientos y resultados inferenciales existe una relación explicativa. Hay dos posibles *explananda*:

- (i) *Pre-data*: propiedades genéricas acerca de cómo pueden errar los procedimientos inferenciales vinculados a un modelo estadístico (e.g., insesgadez)<sup>7</sup>.
- (ii) *Post-data*<sup>8</sup>: propiedades específicas dependientes de una muestra particular (e.g. la varianza de un estimador). El *explanans* estará constituido exclusivamente por las propiedades del mecanismo generador de datos representadas por el modelo estadístico y sus supuestos en el primer caso y, adicionalmente, por el tamaño y valores de los datos particulares, en el segundo caso.

## 4. Errores en estadística y econometría

El foco del trabajo estará puesto en los errores que se producen cuando se hace inferencia estadística. En la inferencia estadística se utiliza la información de una muestra constituida por un subconjunto de los individuos de una población para hacer afirmaciones acerca de características de esa población. Las características más importantes de una población se denominan parámetros, y en la estadística clásica se suponen con un valor determinado pero desconocido. La inferencia

---

<sup>6</sup> El  $p$ -valor es una característica de error *post-data*: una vez obtenida una muestra y el valor específico del estadístico de prueba que es función de esa muestra, el  $p$ -valor es la probabilidad de que la hipótesis nula sea cierta dado el valor específico de ese estadístico de prueba.

<sup>7</sup> El error de una estimación particular no se puede observar, puesto que hacerlo requiere el conocimiento del parámetro estimado, lo cual haría innecesaria la inferencia estadística en primer lugar. En contraste, los enfoques bayesianos le asignan probabilidades a las hipótesis antes y después de la realización de los datos.

<sup>8</sup> Véase Mayo y Spanos (2006) para una discusión acerca de la evaluación *pre-data* y *post-data* de las pruebas de hipótesis e intervalos de confianza.

estadística produce procedimientos para estimar o contrastar hipótesis acerca de esos parámetros a partir de la muestra, y los conceptos de error vinculados a la misma refieren a las maneras y cuantías en que esos procedimientos pueden equivocarse respecto a los verdaderos parámetros. Sin embargo, la posibilidad de errar acompaña a cualquier fenómeno estocástico. En la siguiente subsección ilustraremos los conceptos de error vinculados a variables aleatorias para, a continuación, abordar ejemplos de formas de error en la inferencia estadística.

#### 4.1. El error y los fenómenos estocásticos

Un fenómeno puede ser modelado a través de variables aleatorias. Una variable aleatoria es una función que le asigna un valor numérico a cada uno de los resultados del experimento aleatorio subyacente. Que tratemos a un fenómeno como un experimento aleatorio implica que, antes de su realización, no hay manera de saber el resultado exacto que se obtendrá. Esto introduce una fuente ineludible de error en cualquier conjetura que hagamos respecto al resultado de una variable aleatoria. Sin embargo, el comportamiento aleatorio de una variable puede ser comprendido mediante un modelo que le asigne una distribución de probabilidad o, al menos, algunos de sus momentos. Esto permite una caracterización más específica de los errores.

En particular, la esperanza (o valor esperado) y la desviación estándar constituyen parámetros que facilitan el entendimiento del tipo de error que cometeremos, anticipando el resultado de una variable aleatoria. Si tuviéramos que apostar a un valor específico que tomará la variable aleatoria, la esperanza es el valor que reduciría la distancia al cuadrado del error resultante de repetir el experimento aleatorio infinitas veces. Por otra parte, el desvío estándar es una medida de la dispersión de la variable aleatoria y nos dice, en promedio, cuán lejos estarán los valores de dicha variable respecto de la esperanza. Es una manera de cuantificar el error promedio que cometeremos cuando ‘apostamos’ a la esperanza como el valor que tomará la variable aleatoria y, por ende, una medida de la incertidumbre e imprecisión asociada a esa variable<sup>9</sup>.

Una idea más concreta de los conceptos de error la brindará el análisis de un modelo específico, como el modelo lineal. En una regresión lineal existen tres tipos de variables aleatorias: un conjunto  $X = \{X_1, X_2, \dots, X_k\}$  de variables independientes o explicativas, una variable  $Y$  dependiente o explicada y un término del error  $U$ . En el modelo clásico lineal estas variables guardan la siguiente relación:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + U$$

El modelo lineal puede interpretarse como la respuesta a la siguiente pregunta: si conocemos los valores de los coeficientes (representados por  $\beta_0, \beta_1, \dots, \beta_k$  y que de aquí en adelante llamaremos coeficientes  $\beta$ ) y se realizan valores particulares de las variables independientes, ¿qué valor tomará la variable dependiente? El modelo responde que el valor predicho de la variable dependiente,  $\hat{Y}$ , es  $E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ . El término del error  $U$  (bajo la interpretación habitual) es una variable aleatoria que representa a todos los factores que determinan el valor de  $Y$ , además de las variables dependientes observadas  $X$ , pero que no podemos observar (ni conocer sus

<sup>9</sup> Gauss (1995) propone la varianza como medida para cuantificar la incertidumbre de las observaciones de errores de medición. La desviación estándar es la raíz cuadrada de la varianza.

coeficientes). De esta manera,  $U$  sirve para representar el error que cometeremos al intentar predecir el valor de  $Y$  conociendo los coeficientes de las variables observadas y un valor particular de las mismas. La esperanza y la desviación estándar de  $U$  son, pues, características del error que cometeremos cuando utilizamos el modelo para predecir el valor de  $Y$ , aún si conocemos el verdadero valor de los coeficientes de regresión.

#### 4.2. El error en la estadística inferencial

En la sección anterior sostuvimos que de la especificación de un modelo se siguen propiedades de los errores de los procedimientos inferenciales. Estos procedimientos son, por ejemplo, estimadores, predictores, intervalos de confianza y pruebas de hipótesis. En esta subsección nos enfocaremos en los estimadores y en la que sigue diremos algo acerca de las pruebas de hipótesis.

Un estimador es una regla que asigna un valor estimado de un parámetro a cada posible muestra aleatoria. Como tal, un estimador es una variable aleatoria que es realizada en función de distintas muestras particulares. Las preguntas vinculadas a estimadores que nos interesan son ¿de qué manera podemos equivocarnos al estimar un parámetro? ¿qué causará esos posibles errores? ¿cuándo y cómo podemos evitarlos?

En el caso del modelo lineal, los parámetros de interés son los coeficientes  $\beta$  de la regresión. Es de esperar —dada la diversidad de las muestras posibles y su no necesaria coincidencia con la población— que los estimadores arrojen estimaciones que no coinciden con el parámetro poblacional. El error en la inferencia estadística es inevitable. Sin embargo, los estimadores pueden equivocarse de maneras particulares y esas particularidades configuran propiedades de los estimadores, esto es, características del error que se puede cometer estimando un parámetro mediante un determinado estimador.

De especial importancia son las propiedades de insesgadez, consistencia y eficiencia. Un estimador  $\hat{\beta}^{10}$  de un parámetro  $\beta$  es insesgado cuando la esperanza del estimador es igual al valor del parámetro. Si tomamos distintas muestras de una población y aplicamos un estimador, los valores particulares de la estimación variarán y diferirán en general respecto del parámetro. Sin embargo, cuando el estimador es insesgado la esperanza de la variable aleatoria es igual al parámetro. Un estimador sesgado tendrá una esperanza distinta al parámetro poblacional, lo cual es un error poco tolerable, y de ahí viene la importancia que se le asigna a la insesgadez de un estimador<sup>11</sup>. En el contexto del modelo de regresión lineal, el estimador estándar es el de Mínimos Cuadrados Ordinarios (MCO). La insesgadez de este estimador se garantiza si se cumplen los siguientes cuatro supuestos (Wooldridge, 2015):

---

<sup>10</sup> Nótese el acento sombrero que acompaña a la letra griega y que se usa para referirse a un *estimador* del verdadero parámetro  $\beta$

<sup>11</sup> En muchos casos donde la insesgadez no puede asegurarse, a un estimador se le exige consistencia. La consistencia de un estimador se da cuando la probabilidad de que el valor estimado y el valor del parámetro coincidan tiende a uno a medida que la muestra tiende a infinito

1. **Linealidad en los parámetros:** el modelo poblacional puede escribirse:  
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$
2. **Muestreo aleatorio:** se tiene una muestra aleatoria de  $n$  observaciones,  $\{(x_{i1}, x_{i2}, \dots, x_{ik}, y_i) : i = 1, \dots, n\}$  que sigue el modelo del supuesto 1.
3. **Ausencia de colinealidad perfecta:** hay variación en todas las variables independientes y no hay relación lineal exacta entre ellas.
4. **Media condicional cero:**  $E(u|x_1, x_2, \dots, x_k) = 0$

Un estimador, como toda variable aleatoria, tiene una varianza, la cual es una medida de su imprecisión (cuán diferentes de la esperanza son los valores estimados en promedio). Distintos estimadores pueden tener distintas varianzas. Un estimador es eficiente respecto a otro cuando tiene menor varianza y la eficiencia dentro de alguna clase de estimadores es una propiedad importante o, a la inversa, la imprecisión es un aspecto importante de los errores propios de la inferencia estadística. Si a los cuatro supuestos mencionados se le agrega el siguiente:

5. **Homocedasticidad:**  $var(u|x_1, \dots, x_k) = \sigma^2$

se puede probar el teorema de Gauss-Markov, el cual afirma que el estimador MCO es el estimador lineal insesgado más eficiente o de varianza mínima. La eficiencia es una característica importante de la forma en que los estimadores pueden errar

### 4.3. Explicación, eliminación y tolerancia del error

La demostración de la insesgadez del estimador MCO y el teorema de Gauss-Markov son importantes para nuestros objetivos. Constituyen un marco para la explicación de distintas formas en que es posible equivocarse al utilizar el estimador MCO. Si se cumplieran los cinco supuestos, el error que cometeremos será insesgado y eficiente. Por otro lado, si alguno de esos supuestos fallara, se cometerán errores de otro tipo, más graves (un sesgo, inconsistencia o ineficiencia). Otro lugar en donde se puede ver el potencial explicativo del error de este modelo es en la fórmula de la varianza. Si se cumplen los cinco supuestos, la varianza del estimador MCO  $\hat{\beta}_j$  de la variable independiente  $x_j$  está dada por la siguiente fórmula

$$var(\hat{\beta}_j) = \frac{\sigma^2}{SCT_j(1 - R_j^2)}$$

Donde  $\sigma^2$  es la varianza del error  $U$ ,  $SCT_j$  es la suma de cuadrados totales de  $x_j$  o su variación muestral total  $SCT_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$  y  $R_j^2$  mide la relación que hay entre la variación de  $x_j$  y la variación de las restantes variables independientes. Esta fórmula constituye un ejemplo claro de un vehículo de explicación de la imprecisión. Ella establece que la varianza del término del error hace más impreciso al estimador, que el tamaño y variación de la muestra lo hace más preciso y que, mientras mayor sea la relación de la variable independiente  $x_j$  con las restantes variables independientes, mayor será la imprecisión del estimador.

Otro caso de explicación de errores se da en torno a los posibles sesgos. Los modelos estadísticos proveen la posibilidad de explicar la producción de sesgos en los estimadores. Fuentes de sesgos en los estimadores son, por ejemplo, la omisión de variables o los errores de medición. Por la importancia que tiene en el desarrollo de técnicas para evitarlo, vale la pena examinar el sesgo de omisión de variables, el cual se produce porque la omisión de una variable relevante lleva a veces a que no se cumpla el cuarto supuesto de Gauss-Markov (generando ‘endogeneidad’). Suponiendo que el modelo adecuado (que cumple con los cuatro primeros supuestos de Gauss-Markov) es:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$

Pero estimamos un modelo que omita  $X_2$ :

$$Y = \widetilde{\beta}_0 + \widetilde{\beta}_1 X_1 + E$$

No es difícil demostrar que  $E(\widetilde{\beta}_1) = \beta_1 + \beta_2 \delta_1$ , donde  $\delta_1$  es el coeficiente de una regresión auxiliar de  $X_2$  sobre  $X_1$  y  $\beta_2$  es el coeficiente de la regresión de  $Y$  sobre  $X_2$  en el verdadero modelo. De esta manera,  $\beta_2 \delta_1$  es el sesgo que se produce por omitir  $X_2$  y esta fórmula permite explicar, por ejemplo, el signo y la magnitud del sesgo. Si  $\beta_2$  y  $\delta_1$  tienen el mismo signo, el sesgo será positivo, en caso contrario será negativo. La omisión de variables implica que no es posible estimar ninguno de los dos, pero distintas hipótesis acerca de su valor habilitan distintas hipótesis acerca de la dirección y magnitud del sesgo. La endogeneidad no solo provoca que el estimador MCO sea sesgado, sino que también sea inconsistente, lo cual es un problema grave.

Los modelos no solo permiten caracterizar la omisión de variables y explicar el sesgo que producen, sino que también plantear estrategias para evitar el error asociado al sesgo. El uso de variables instrumentales y de variables *proxies*<sup>12</sup> son algunos ejemplos de las soluciones posibles a este problema. Lo relevante para los objetivos de este trabajo es que estas soluciones surgen en conexión con la explicación del sesgo que se produce al omitir variables y que se plantean en el marco de modelos apropiados. Por ejemplo, cuando hay omisión de variables y violación del cuarto supuesto de Gauss-Markov, el estimador de variables instrumentales es consistente siempre que se cumplan determinados supuestos alternativos<sup>13</sup>. De esta manera, la violación de un supuesto de Gauss-Markov que provoca que el estimador MCO cometa errores intolerables es parcialmente reparado con un modelo y un estimador alternativos que deben cumplir con otros supuestos para que sean apropiados.

Para finalizar esta sección es pertinente ilustrar el uso de modelos estadísticos para tolerar los errores que no se pueden evitar. Este uso es patente en las pruebas de hipótesis, en donde se parte por plantear una hipótesis nula  $H_0$  y una hipótesis alternativa  $H_1$  respecto del valor que asume un

<sup>12</sup> Una variable *proxy* es una variable que se usa en reemplazo de la variable omitida. Debe guardar una correlación con la variable omitida que puede no ser perfecta. Por ejemplo, el PIB per cápita puede usarse como variable *proxy* del nivel de vida de un país.

<sup>13</sup> Estos son el supuesto de exogeneidad (i.e. la variable instrumental afecta a la variable dependiente del modelo original solo a través de la variable endógena) y el supuesto de relevancia (i.e. la variable instrumental está altamente correlacionada con la variable endógena).

parámetro. Siendo ambas hipótesis exhaustivas y excluyentes, las pruebas de hipótesis construyen un estadístico de prueba concebido para cuantificar la cercanía o lejanía de los datos respecto de la hipótesis nula y reglas de decisión según las cuales se rechaza o no se rechaza la hipótesis nula de acuerdo con el valor que adopte el estadístico de prueba. Al usar estas reglas, se pueden cometer dos tipos de errores: el error de tipo I o falso positivo, en el cual se rechaza la hipótesis nula cuando es verdadera, y el error de tipo II o falso negativo en el cual no se rechaza la hipótesis nula cuando es falsa. Supuestos adicionales acerca del modelo original permiten deducir la distribución que seguirá el estadístico de prueba. En el caso del modelo lineal, un supuesto de este tipo es el siguiente:

#### 6. Normalidad del error: $u \sim Normal(0, \sigma^2)$

Bajo este supuesto, el estadístico de prueba de una prueba de significancia individual de un parámetro sigue una distribución t de *Student* con  $n$  grados de libertad. Esta información permite cuantificar la probabilidad de que el estadístico de prueba tome un rango de valores suponiendo que la hipótesis nula es verdadera. En el caso de la prueba de significancia individual el modelo muestra que, cualquiera sea el valor crítico del estadístico de prueba que nuestra regla establezca para rechazar la hipótesis nula, siempre habrá alguna probabilidad de cometer un error de tipo I. El modelo permite explicar las probabilidades de error de tipo I y de tipo II asociadas a distintas reglas de decisión. En particular, da cuenta de la relación inversa entre ambas probabilidades: mientras más pequeña queramos que sea la probabilidad de cometer un error de tipo I, más error de tipo II tendremos que estar dispuestos a tolerar. Por lo general, se elige la probabilidad de cometer un error de tipo I (nivel de significancia) del 1%, 5% o 10% y se asume lo que esto conlleva respecto a las probabilidades de cometer un error de tipo II.

Para los propósitos de este trabajo, cabe destacar que los supuestos acerca del mecanismo estocástico permiten explicar características y relaciones acerca de los errores realizables en el contexto de una prueba de hipótesis. En particular, muestran el carácter inevitable del error de tipo I y II y su relación inversa. Dado el carácter inevitable de estos errores, estas explicaciones hacen posible implementar criterios de tolerancia a través de las reglas de decisión para rechazar o no rechazar la hipótesis nula. Por lo tanto, los modelos y sus supuestos facilitan explicaciones del error que a su vez ayudan a gestionar su tolerancia.

## 5. Modelos y explicación

En esta sección se especificará en qué sentido los modelos estadísticos cumplen una función explicativa del error. La literatura acerca de modelos y explicación está enfocada en cómo los modelos teóricos<sup>14</sup> proveen explicaciones de fenómenos sustantivos. Plantearemos continuidades y diferencias de esta literatura con la relación entre las explicaciones y los modelos estadísticos.

<sup>14</sup> Existe una amplia literatura que discute qué son los modelos teóricos y no es la intención de este trabajo tomar una postura al respecto. De manera laxa y abarcadora se puede decir que en economía los modelos teóricos son modelos que buscan reproducir algún hecho estilizado, explicarlo, hacer predicciones sobre un fenómeno o representarlo, sin la pretensión de representar un mecanismo estocástico que produzca los datos utilizados para estimar los parámetros del modelo.

Bokulich (2011, pp.38-39) plantea una concepción acerca de modelos y explicaciones basada en tres nociones:

1. El *explanans* refiere a un modelo científico que contiene idealizaciones o ficcionalizaciones
2. Siguiendo a Woodward (2003, p.11) el modelo explica en virtud de proveer información acerca de patrones de dependencia contrafáctica.
3. El modelo se justifica como un buen modelo en relación con las preguntas que pretende responder y es apropiado a su dominio de aplicación.

A continuación, veremos cómo esta concepción se aplica a la explicación del error a través de modelos estadísticos que hemos desarrollado en las secciones anteriores. En la siguiente subsección se analizará el segundo requisito, para luego hacerlo con el primero y el tercero.

### 5.1. Modelos y dependencia contrafáctica

La concepción de la explicación científica de Woodward es una concepción causal. Vale la pena citar un pasaje en el que explicita su postura.

(...) se debe poder asociar a toda explicación exitosa un experimento hipotético y contrafactual que nos muestra que la manipulación de los factores mencionados en la explicación (el *explanans*, como lo llaman los filósofos) sería una manera de manipular o alterar el fenómeno explicado (el *explanandum*) y cómo lo hace. Puesto de otra forma, una explicación debe ser tal que permita responder lo que llamo preguntas del tipo *qué pasaría-si-las-cosas-fueran-diferentes*: la explicación debe habilitarnos mirar qué clase de diferencia se hubiera producido en el *explanandum* si los factores citados en el *explanans* hubieran sido diferentes en varias maneras posibles. Podemos pensar esto como información acerca de un patrón de dependencia contrafáctica entre el *explanans* y el *explanandum* [...] (Woodward, 2003, p.11)

En el caso de los modelos estadísticos, lo que cumple el rol de *explanandum* son las propiedades genéricas de error (*pre-data*) de los procedimientos inferenciales o las propiedades específicas (*post-data*) de los resultados inferenciales. Aquello que cumple el rol de *explanans* son, por un lado, los supuestos hechos sobre el modelo estadístico (e.g. los supuestos de Gauss-Markov) y, por otro lado, parámetros específicos de la población o aspectos específicos de los datos (e.g. el tamaño y variación de la muestra, la varianza del error, u otros). Las intervenciones hipotéticas sobre el *explanans* son manipulaciones que imprimen determinadas características en el mecanismo estocástico mientras las restantes se mantienen fijas. Los modelos estadísticos a través de relaciones deductivas habilitan distintos tipos de preguntas del tipo ‘qué pasaría si este supuesto o parámetro o tamaño de muestra fuera distinto’.

Volviendo al ejemplo de la varianza de un estimador MCO, la derivación de su fórmula depende del cumplimiento de los supuestos de Gauss-Markov y los valores que toma depende de la varianza del término de error, el tamaño y variación de la muestra y de la relación con otras variables independientes. Es claro que la fórmula provee información acerca de los patrones de dependencia

contrafáctica entre estos factores y la varianza del estimador, dando cuenta de cómo posibles manipulaciones de esos factores cambiarían la precisión del estimador.

Otro ejemplo vinculado a manipulaciones efectivas y no solo hipotéticas son los análisis del poder de una prueba de hipótesis. El poder de una determinada prueba es la probabilidad que tiene, dado un nivel de significancia, de detectar una discrepancia de la hipótesis nula. Mientras mayor sea el poder de una prueba de hipótesis, más probabilidad hay de declarar un efecto estadísticamente significativo cuando existe. El poder depende, entre otras cosas del tamaño de la muestra, por lo cual en la investigación empírica a veces se elige el tamaño de la muestra con el objetivo de alcanzar un determinado poder. El poder es igual a  $1 - \beta$  donde  $\beta$  es la probabilidad de cometer un error de tipo II (no rechazar la hipótesis nula cuando es falsa). Por lo tanto, es claramente uno de los *explananda* que se abordan en este trabajo.

Los patrones de dependencia contrafáctica más relevantes que son informados por los modelos estadísticos son los que se dan entre los supuestos de un modelo y las propiedades genéricas de error de los procedimientos inferenciales. Es claro que teoremas como los que exhiben la insesgadez, consistencia o eficiencia de un estimador generan esta información, puesto que responsabilizan causalmente a características del mecanismo generador de datos (representadas en los supuestos del modelo) respecto de las formas en que pueden errar los procedimientos inferenciales. Es patente, además, cómo la estadística y econometría se han desarrollado respondiendo preguntas del tipo qué-pasaría-si-estos-supuestos-fueran-diferentes. Las alteraciones de los supuestos y el planteo de modelos alternativos pueden concebirse como intervenciones hipotéticas<sup>15</sup> sobre un mecanismo generador de datos para ver su efecto en la caracterización del error de los procedimientos inferenciales. Entonces, se puede decir legítimamente que los modelos estadísticos facilitan la explicación de cómo y por qué los procedimientos inferenciales yerran.

## 5.2. Explicación y eliminación del error

Una tesis de este trabajo es que la explicación del error en los modelos estadísticos es un medio para aminorarlos y tolerarlos donde corresponda. Ahora bien, esto se relaciona con la concepción del vínculo entre modelos y explicación que estamos desarrollando de dos maneras: la primera es que cuando se está ante un error intolerable (como la inconsistencia de un estimador) la investigación de los patrones de dependencia contrafáctica que lo producen puede eventualmente encontrar intervenciones que lo eliminen. La concepción intervencionista de la causalidad de Woodward sobre la que subyace su modelo de explicación causal es especialmente afín al descubrimiento de posibles intervenciones que modifiquen el *explanandum*

<sup>15</sup> Las simulaciones computacionales en estadística constituyen intervenciones reales y no hipotéticas sobre un mecanismo generador de datos para ver su efecto en propiedades de estimadores, sesgos, ilustrar teoremas, etc. En ellas se crea un mecanismo generador de datos (a través del uso de números pseudo-aleatorios) al cual se le imprimen determinadas características. Por ejemplo, se puede generar el mecanismo de un modelo lineal con los supuestos de Gauss-Markov en donde se fija el verdadero valor de los parámetros. Luego es posible estudiar el comportamiento del estimador MCO y ver si es compatible con las propiedades demostradas por los teoremas. También es factible estudiar el efecto de la violación de un supuesto. Todas estas son intervenciones reales y no solo hipotéticas sobre un mecanismo generador de datos.

Ya se mencionó cómo el estimador de variables instrumentales surge como solución a los problemas de endogeneidad por variables omitidas en el estimador MCO. La adopción de variables instrumentales puede entenderse como una intervención tendiente a solucionar el problema que genera la endogeneidad de una de las variables independientes en el modelo original. La teoría asociada a las variables instrumentales estudia parte del acervo de relaciones de dependencia contrafáctica, en donde la inconsistencia y el sesgo figuran como *explananda*. Existe en la concepción intervencionista de la causalidad una continuidad entre la información de intervenciones que producen un explanandum y las intervenciones que pueden eliminarlo.

La segunda forma en que la gestión del error se vincula con la concepción de la explicación científica acá adoptada radica en que la explicación más profunda de una determinada forma de errar permite evitar otras formas de errar. Para Woodward (2003, p.260), la profundidad de una explicación viene dada por la apelación a generalizaciones que son invariantes bajo un rango amplio e importante de intervenciones y que pueden responder más preguntas del tipo qué-pasaría-si-las-cosas-fueran-diferentes. Cuando, por ejemplo, se quiere asegurar la consistencia de un estimador, se investiga una relación de dependencia contrafáctica entre propiedades del mecanismo generador de datos y la consistencia del estimador. La violación de un supuesto que provoca la inconsistencia del estimador puede entenderse como una intervención que no deja invariante la relación entre el mecanismo generador de datos y la consistencia del estimador.

Ante esto, la estadística y econometría buscan modelos y estimadores que restauren esta relación como, por ejemplo, el modelo y estimador de variables instrumentales. Si la consistencia es nuestro *explanandum* y el mecanismo generador de datos es nuestro *explanans* estas investigaciones pueden verse como búsquedas de generalizaciones invariantes bajo un rango más amplio de intervenciones. El fenómeno de la consistencia de un estimador es explicado más profundamente. Pero el objetivo primario de ellas no es explicar más profundamente la consistencia, sino que eliminar la inconsistencia cuando se hace presente. Vemos que la explicación más profunda de cómo un estimador puede errar consistentemente resulta un medio para evitar la inconsistencia como error intolerable.

### 5.3. Idealizaciones y modelos estadísticos

Corresponde a continuación examinar el primer y tercer requisito de la explicación a través de modelos que plantea Bokulich (2011). El primero requiere que la explicación refiera a modelos que incluyen idealizaciones o ficcionalizaciones, y los modelos estadísticos incluyen claras idealizaciones. Un modelo que asigna una determinada distribución de probabilidad a un mecanismo generador de datos idealiza algunos aspectos de este. Por ejemplo, modelar las variables aleatorias como continuas y con una función de distribución de probabilidad continua idealiza una variable aleatoria que puede tener saltos discretos y que no se ajusta a la perfección a la distribución modelada. El hecho de que las distribuciones usadas en las pruebas de hipótesis supongan un rango infinito de valores también puede considerarse una idealización respecto de un mecanismo que tiene un rango finito de valores. Supuestos como el de esperanza condicional cero del término del error (i.e. el cuarto de los supuestos de Gauss-Markov en el ejemplo dado más arriba) son aplicables también cuando el mecanismo tiene una media condicional cercana a ese valor.

Las distorsiones propias de un modelo estadístico pueden considerarse representaciones idealizadas donde las tosquedades del mecanismo real son suavizadas para permitir su tratabilidad matemática. Es muy importante notar que estas distorsiones no son ficciones como las que admite Bokulich, y esto marca una importante diferencia entre los modelos estadísticos y los modelos teóricos en su rol explicativo. Sin embargo, no es este el lugar para desarrollar una postura acerca de explicaciones que involucran ficciones. Solo cabe señalar que, en principio, las ficciones propias de un modelo podrían ser instrumentales para informar relaciones de dependencia contrafáctica. Por ejemplo, un modelo de equilibrio parcial podría suponer una firma representativa que, por sus características, es una ficción. Sin embargo, el modelo permitiría informar de relaciones de dependencia contrafáctica entre cambios en el precio de los factores y la intensidad de los factores. El modelo en este caso es evaluado con base en la adecuación de estas relaciones y no por la adecuación de la ficción que involucra<sup>16</sup>.

¿Por qué no podrían utilizarse ficciones en el contexto de modelos estadísticos? La respuesta a esta pregunta radica en que aquellos aspectos de un mecanismo generador de datos que resultan modelados son los que son identificados como causas del *explanandum*. En el ejemplo que planteábamos previamente, la ficción de la firma representativa no tiene un rol explicativo causal en la proporción de capital y trabajo que adopta un mercado. Ese rol lo cumplen los cambios en los precios relativos de los factores. Esta diferencia le ponen una exigencia a los supuestos de un modelo estadístico que no tienen todos los modelos teóricos.

Esto nos lleva al tercer requisito, a saber, que el modelo esté justificado como un buen modelo. Para explicar, el modelo debe ser adecuado. La adecuación de un modelo estadístico radica en que represente bien las características del mecanismo generador de datos. La especificación de un modelo es la determinación de sus supuestos, las variables que se incluyen o excluyen, las formas funcionales que se adoptan, etc. Las propiedades de los procedimientos inferenciales se siguen deductivamente de los supuestos del modelo y para que esas inferencias sean sólidas los supuestos deben ser verdaderos. Por ello, las pruebas de malas especificaciones juegan un rol central en el uso de modelos para la inferencia estadística (Spanos, 2006b). Los supuestos de un modelo deben ser validados para que su función explicativa e inferencial se realice adecuadamente.

## 6. Conclusiones

El trabajo analizó la función que los modelos estadísticos cumplen en la explicación del error. La concepción intervencionista de la explicación causal se mostró especialmente afín al rol que estas explicaciones tienen en hallar intervenciones que reemplacen formas intolerables de error por formas tolerables.

Concluiremos con una observación acerca de los límites que tienen las explicaciones mediante modelos estadísticos: no existe una sola gran explicación ni un solo gran modelo de cualquier *explanandum* interesante. El error constituye un problema lo suficientemente exuberante para que ningún modelo estadístico alcance a explicar todas las maneras en las que podemos equivocarnos

---

<sup>16</sup> Se trata de una idea que está presente en el trabajo de Friedman (1953). Aunque se lo suele interpretar de manera instrumentalista (Caldwell, 1980) dejando afuera a la explicación como fin epistémico, es posible compatibilizar su postura con la explicación como fin.

cuando hacemos investigación empírica. El desarrollo de una teoría del error es una tarea que no tiene final y constituye el progreso de las disciplinas dedicadas a la investigación empírica. Los modelos son una herramienta fundamental para ello.

## Referencias

- Bacon, F. (2019). *Novum organum; or true suggestions for the interpretation of nature*. Good Press.
- Bokulich, A. (2011). How scientific models can explain. *Synthese*, 180(1), 33–45. <https://doi.org/br8vd2>
- Caldwell, B. J. (1980). A critique of friedman's methodological instrumentalism. *Southern Economic Journal*, 47(2), 366–374. <https://doi.org/d6ssfz>
- Descartes, R. (2013). *Meditations on first philosophy*. Broadview Press.
- Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, 222(594604), 309–368. <https://doi.org/ddntd6>
- Friedman, M. (1953). *Essays in positive economics*. University of Chicago press.
- Gallop, D. (1984). *Parmenides of Elea: A text and translation with an introduction*. University of Toronto Press. <http://www.jstor.org/stable/10.3138/j.ctt1287z63>
- Gauss, C. F. y Stewart, G. W. (1995). *Theory of the combination of observations least subject to errors, part one, part two, supplement*. SIAM. <https://doi.org/10.1137/1.9781611971248>
- Laudan, L. (1984). *Science and values: The aims of science and their role in scientific debate*. University of California Press. <https://doi.org/10.1525/9780520908116>
- Mayo, D. G. (1996). *Error and the growth of experimental knowledge*. University of Chicago Press. <http://dx.doi.org/10.7208/chicago/9780226511993.001.0001>
- Mayo, D. G. y Spanos, A. (2006). Severe testing as a basic concept in a neymanpearson philosophy of induction. *British Journal for the Philosophy of Science*, 57(2), 323–323. <https://doi.org/d62nz5>
- Mayo, D. G., y Spanos, A. (2011). Error statistics. En D. Gabbay, P. Thagard y J. Woods (Eds.), *Philosophy of statistics* (pp.153–198). Elsevier. <https://doi.org/b9wkn3>
- Popper, K. (2014). *Conjectures and refutations: The growth of scientific knowledge*. Routledge. <http://dx.doi.org/10.4324/9780203538074>
- Platón. (2004). *Plato: Republic* (Trad. Reeve et al.). Hackett Publishing.
- Reiss, J. (2008). *Error in economics. towards a more evidence-based methodology*. Routledge. <https://doi.org/c83f3q>

- Spanos, A. (2006a). Econometrics in retrospect and prospect. En T. Mills y K. Patterson (Eds.), *New palgrave handbook of econometrics, vol. 1: Theoretical econometrics* (pp.3– 58). Palgrave Macmillan
- Spanos, A. (2006b). Where do statistical models come from? revisiting the problem of specification. En J. Rojo (Ed.), *Optimality: The second Erich L. Lehmann symposium* (pp.98–119). Institute of Mathematical Statistics. <http://dx.doi.org/10.1214/074921706000000419>
- Spanos, A. (2012). Philosophy of econometrics. En U. Mäki (Ed.), *Philosophy of economics* (pp.329–394). Elsevier. <http://dx.doi.org/10.1016/B978-0-444-51676-3.50013-0>
- Spanos, A. (2021). Philosophy of econometrics. En C. Heilmann y J. Reiss (Eds.), *The Routledge handbook of philosophy of economics* (pp.397–422). Routledge. <http://dx.doi.org/10.4324/9781315739793-36>
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press. <http://dx.doi.org/10.1093/0195155270.001.0001>
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.
- Spanos, A. (2006a). Econometrics in retrospect and prospect. En T. Mills y K. Patterson (Eds.), *New palgrave handbook of econometrics, vol. 1: Theoretical econometrics* (pp.3– 58). Palgrave Macmillan
- Spanos, A. (2006b). Where do statistical models come from? revisiting the problem of specification. En J. Rojo (Ed.), *Optimality: The second Erich L. Lehmann symposium* (pp.98–119). Institute of Mathematical Statistics. <http://dx.doi.org/10.1214/074921706000000419>
- Spanos, A. (2012). Philosophy of econometrics. En U. Mäki (Ed.), *Philosophy of economics* (pp.329–394). Elsevier. <http://dx.doi.org/10.1016/B978-0-444-51676-3.50013-0>
- Spanos, A. (2021). Philosophy of econometrics. En C. Heilmann y J. Reiss (Eds.), *The Routledge handbook of philosophy of economics* (pp.397–422). Routledge. <http://dx.doi.org/10.4324/9781315739793-36>
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press. <http://dx.doi.org/10.1093/0195155270.001.0001>
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.