ASOCIACION ARGENTINA
DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

# LIII Reunión Anual

Noviembre de 2018

ISSN 1852-0022
ISBN 978-987-28590-6-0

## Fundamentos probabilísticos de la econometría

**González Mirta Lidia**
**Landro Alberto Héctor**

# Fundamentos probabilísticos de la econometría

***Mirta L. González***
***Alberto H. Landro***
*Centro de Investigaciones en Econometría*
*IADCOM-FCE-UBA*

## Resumen

Contrariamente a lo afirmado por la tradición literaria de la econometría, se demuestra que la génesis de la teoría de modelos estocásticos, cuya culminación se encuentra en el criterio de los cuadrados mínimos y en la síntesis de Gauss-Laplace, se remonta al problema de la inversión de la probabilidad y la propuesta de solución de Simpson, dirigida al análisis de los errores aleatorios. La generalización de este criterio de optimización basado en la minimización de la suma de los cuadrados de los errores y la asimilación de los fenómenos dinámicos a procesos estocásticos dieron origen a la teoría de los modelos econométricos, que evolucionaron de una posición inicial determinística hacia el teorema de descomposición predictiva de Wold.

Palabras clave: modelos econométricos, probabilidad

Códigos de clasificación JEL: C51, C01, B16, B23

## Abstract

Contrary to what is affirmed by the literary tradition of econometrics it is demonstrated that the genesis of the theory of stochastic models, whose culmination is found in the criterion of minimum squares and in the synthesis of Gauss-Laplace, goes back to the problem of inversion of the probability and the solution proposal of Simpson directed to the analysis of the random errors. The generalization of this optimization criterion based on the minimization of the sum of the squares of the errors and the assimilation of the dynamic phenomena to stochastic processes gave rise to the theory of the econometric models, which evolved from a deterministic initial position to the theorem of predictive decomposition of Wold.

Keywords: Econometric models, probability

JEL Codes: C51, C01, B16, B23

## 1.- El problema de la inversión de la probabilidad y la ley de los grandes números

En 1713, como recopilación de sus trabajos inéditos sobre teoría de la probabilidad de los 20 años anteriores, se publicó el "*Ars conjectandi*" de Jakob Bernoulli. Una obra que, por su tratamiento absolutamente novedoso de la evidencia empírica, modificó los fundamentos de la probabilidad en tal forma que debería ser considerada como el verdadero punto de partida de la teoría de la probabilidad y la culminación del proceso de formación del concepto de probabilidad.

La contribución fundamental de J. Bernoulli -incluida en la "*Pars quarta, Fradens usum et applicationem procedentis doctrinæ in civilibus, moralibus et œconomicis"* del "*Ars conjectandi"*- consistió en: **i)** proporcionar la primera definición estricta de probabilidad en un contexto matemático como "*...grado de certeza (gradus certitudinis), que difiere de ésta de la misma forma que una parte con respecto al todo"*[1], **ii)** postular que se podía "*aprender"* a partir de la experiencia, que cuanto mayor fuera la cantidad de información con que contara el observador más aproximada sería la estimación de su verdadero valor; **iii)** demostrar que este aprendizaje era cuantificable a través de un proceso de inversión de la probabilidad y, en consecuencia, asumiendo las hipótesis de independencia y regularidad, **iii)** establecer el nexo entre las probabilidades "a priori" (definidas a partir de un razonamiento que va de las causas a los efectos, de la simetría de los resultados posibles al concepto de equiprobabilidad) y las probabilidades "a posteriori" (definidas a partir de un razonamiento que va de los efectos a las causas.

Esta demostración del principio intuitivo que la incertidumbre disminuía en la medida que se incrementaba el número de observaciones repetidas en igualdad de condiciones, y la cuantificación de dicho proceso de inferencia -conocida luego como la "*ley (débil) de los grandes números"*-, constituyó el primer teorema límite de la teoría de la probabilidad[2].

La postura (más teológica y filosófica que matemática) de convencido militante del determinismo metafísico condujo a Bernoulli a la identificación de las causas ignoradas del comportamiento de los fenómenos con un parámetro $\theta$, determinado e invariable y representativo de una naturaleza gobernada por leyes inmutables y, en consecuencia, a una interpretación de su teorema que consideraba a la probabilidad $\theta$ como conocida, limitando sus alcances a la demostración de la convergencia en-probabilidad de la frecuencia relativa $Y_n$ a $\theta$.

A partir de esta demostración Bernoulli intentó una extensión de su teorema -de importancia fundamental en la teoría de la inferencia- que implicaba un planteo inverso de la forma: si se observa que la frecuencia relativa " *converge a un valor determinado*", $\theta$, entonces este valor definirá la "*ley*" que gobierna a dicho evento. Pero la circularidad de este esquema de razonamiento en el cual la convergencia en-probabilidad de las frecuencias relativas se

[1] En *"Die Werke von Jakob Bernoulli"*, vol. 3 (p. 239).

[2] La denominación de "*ley de los grandes números*" referida a una generalización de este teorema relacionada con el comportamiento de los promedios es muy posterior y se debe a Poisson (1837). La forma más rudimentaria de la ley de los grandes números es atribuible a Girolamo Cardano ("*Practica aritmeticæ et mensurandi singularis"* (1539)), quien demostró que el número de ocurrencias de un resultado, en una "*larga serie"* de repeticiones independientes de un fenómeno de comportamiento eventual ($X$), es aproximadamente igual a $n\theta$ (donde $\theta$ denota la probabilidad "a priori" de ocurrencia de dicho resultado en una repetición determinada). Cardano –quien puede ser considerado como el primer gran aleatorista- consideró que la aproximación obtenida al cabo de infinitas repeticiones estaba sujeta a las "*perturbaciones de la fortuna"* (la "*suerte"* en su nomenclatura). Su metafísica podría ser resumida en la siguiente expresión, debida a Daston (1988): "*...el cálculo no puede gobernar lo contingente"* (p. 36). Una idea similar sobre el tratamiento cualitativo del aprendizaje a partir de la experiencia puede hallarse en Halley ("*An estimate the degrees of mortality of mankind drawn from the curious tables of the births and funerals at the city of Breslau; with an attempt to ascertain the price of annuities upon lives"* (1693)), en "*De incerti æstimatione"* de Leibniz (1678) y, como se vio más arriba, en "*La logique, ou l'art de penser"* (1662) de Arnauld-Nicole.

verificaba porque los eventos estaban regidos por una ley determinada pero, a su vez, la convicción de que los eventos se regían por una ley determinada se fundaba en el postulado de inversión de la probabilidad según el cual las frecuencias relativas debían converger a $\theta$, no le permitió obtener una justificación rigurosa de los alcances de su teorema como fundamento de una teoría de la inversión de la probabilidad.

## 2.- Thomas Simpson y el análisis de los errores de observación

Los primeros intentos –fallidos- de solución a las limitaciones inherentes a los postulados de la versión inversa del teorema de Bernoulli se debieron a su sobrino Nikolaus (1709) y a Abraham de Moivre (1718)(1730)(1733).

El análisis de sus trabajos permite concluir que los fracasos de los Bernoulli y de Moivre son atribuibles a su fidelidad a la teología Newtoniana, que les hacía ver en sus versiones de los teoremas límite, el argumento según el cual la estabilidad de las frecuencias estadísticas demostraba la presencia de la "*Divina Providencia*"[3].

Uno de los autores que traicionó ese precepto fue Simpson (1755). Su trabajo se orientó al estudio de las propiedades de los errores de observación no sistemáticos (aleatorios) en astronomía. En este sentido, propuso una distribución triangular continua, demostró que, de acuerdo con esta distribución, el valor medio estaba menos afectado por dichos errores no-sistemáticos que cualquier observación individual e intentó hallar su distribución en el límite.

Los problemas planteados por la astronomía condujeron gradualmente al principio de la media aritmética -calculada a partir de observaciones realizadas en igualdad de condiciones- y a distintos métodos de estimación basados en modelos paramétricos, cuya culminación fue el método de los cuadrados mínimos. Pero, como se verá más adelante, la teoría de los errores de observación, que constituyó su fundamento, recién surgió con un tratamiento sistemático en el siglo XVIII, en "*Opera miscellanea sive æstimatio errorum in mixta mathesiper variationes partium trianguli plani et sphærici*" (1722) de Roger Côtes.

En estas memorias demostró, en primer lugar, que supuesta una hipótesis acerca del carácter probabilístico de las observaciones sobre un fenómeno $Y$, el valor promedio de los valores observados era superior, como estimador ($\hat{Y}$) del verdadero valor de $Y$, a cualquier observación individual: "*Al tomar el valor promedio de un número de observaciones, disminuye en forma apreciable la probabilidad de cometer errores pequeños y prácticamente hace desaparecer toda probabilidad de cometer errores de gran magnitud. Esta última consideración, por sí misma, parece suficiente para hacer recomendable el uso de este método, no sólo a los astrónomos, sino a todos los que estén relacionados con cualquier tipo de información obtenida experimentalmente. Y cuanto mayor sea el número de observaciones realizadas, menor será la propensión a cometer errores, suponiendo que las observaciones se realicen en igualdad de condiciones*"[4].Esta versión de la ley de los grandes números de Simpson no aparece sustentada por ninguna demostración teórica.

Como se mencionó más arriba, la novedad conceptual introducida por Simpson consistió en dirigir la atención sobre los errores aleatorios (es decir, sobre las diferencias entre las observaciones registradas y el "verdadero valor" del fenómeno)[5] -asumiendo una hipótesis

[3] Pearson, K. (1978) "*Los matemáticos ingleses post-Newtonianos experimentaron una mayor influencia de la teología Newtoniana que de su matemática*" (p. 202).
[4] Simpson (1755, pp. 92-93).
[5] En realidad, fue Galileo Galilei ("*Dialogo sui massimi sistemi del mondo*" (1632)) quien introdujo en la literatura el problema de los errores de observación. Si bien no obtuvo una solución cuantitativa ni analítica, arribó a ciertas conclusiones que interpretaban profundamente la esencia del problema: que dichos errores son

específica acerca de su distribución de probabilidades- y no sobre las observaciones en sí mismas, ni sobre los objetos observados[6]. Este planteo, que le permitió -aún sin poseer una teoría de la inferencia- medir el error de estimación (es decir, la incertidumbre de la predicción en términos de una distribución de probabilidades dada), constituyó indudablemente el avance fundamental de su trabajo respecto al de de Moivre.

El principio planteado por Simpson implicaba que, una vez asignada una distribución de probabilidades a los errores de observación, la inversión de la probabilidad debía darse por añadidura, postulando que la distribución de probabilidades de $Y$ condicionada por $\hat{Y}$, debía ser tal que $p(Y / \hat{Y}) \propto p(\hat{Y} / Y)$[7]. En particular, en su publicación de 1755, Simpson supuso que, dado un conjunto de $n$ observaciones independientes, los errores contenidos en cada una de ellas ($\varepsilon_i = Y_i - \widehat{Y_i}, i = 1,2,\dots,n$) podían asumir los valores:

$$-\varepsilon, -\varepsilon + 1, \dots, -2, -1, 0, 1, 2, \dots, \varepsilon - 1, \varepsilon$$

con probabilidades proporcionales a la distribución:

$$q^{-\varepsilon}, q^{-\varepsilon+1}, \dots, q^{-2}, q^{-1}, q^0, q^1, q^2, \dots, q^{\varepsilon-1}, q^{\varepsilon}$$

o a la distribución:

$$q^{-\varepsilon}, 2q^{-\varepsilon+1}, \dots, (\varepsilon - 1)q^{-2}, \varepsilon q^{-1}, (\varepsilon + 1)q^0, \varepsilon q^1, (\varepsilon + 1)q^2, \dots, 2q^{\varepsilon-1}, q^{\varepsilon}$$

(quedando definida una distribución de probabilidades para cada $q > 0$)[8].

Simpson desarrolló solamente el caso para $q = 1$, obteniendo una distribución simétrica de los errores. Si bien la selección de las distribuciones de probabilidades de los errores fue tomada de los trabajos de de Moivre, la forma de su utilización fue totalmente novedosa; adelantándose a Lagrange y a Laplace en más de diez años, desarrolló expresiones para la probabilidad de que el valor medio de los errores del conjunto de observaciones no superara una cantidad dada, y comparó esta probabilidad con las correspondientes a las observaciones individuales.

Para la selección de la primera distribución de probabilidades se basó, en particular, en la solución de de Moivre al problema de la definición de la variable aleatoria que representa la suma de los resultados de $n$ tiradas de un dado "clásico" con $k$ caras, numeradas de 1 a $k$, cuyas probabilidades están dadas por los coeficientes del desarrollo de la expresión:

$$\left(1 + q + q^2 + q^3 + \cdots + q^{k-1}\right)^n = \frac{\left(1 - q^k\right)^n}{(1 - q)^n}$$

---

inevitables, que se distribuyen simétricamente, que la probabilidad de cometer un error de medición aumenta a medida que disminuye la medida del error y que la mayoría de las observaciones se agrupan alrededor del "verdadero valor". Obsérvese que, en estos postulados, están contenidas las características de la distribución Normal que, con el tiempo, se constituiría en una de las funciones básicas de la teoría de la probabilidad y de los rudimentos de la teoría de la contrastación de hipótesis. Conclusiones similares recién pueden ser halladas en "*Beyträge zum Gebrauche der Mathematik und deren Anwendung"* (1772) de J. H. Lambert.

[6] Simpson supuso que la distribución de probabilidades de los errores era siempre conocida, aún cuando la *"verdadera posición"* del objeto observado pudiera considerarse desconocida (debe recordarse que Simpson era astrónomo). Esta hipótesis constituyó el elemento fundamental para la justificación de cualquier método referido a la estimación mediante la utilización de conjuntos finitos de observaciones.

[7] Stigler, S. M. (1986a): "*Simpson había comprendido que el concepto de distribución de los errores permitía un acceso a la medida de la incertidumbre por la puerta de servicio. Fue Laplace quien, colándose posteriormente por ella, logró abrir la puerta principal (sólo para hallar que la llave de Bayes ya se encontraba en la cerradura)"* (p. 98).

[8] Simpson observó que la suma de los términos de la última sucesión elevados a la potencia $t$ es igual a$q^{-\varepsilon t}(1 + q + q^2 + \cdots + q^{\varepsilon})^{2t} = \frac{1-q^{\varepsilon t+1}}{q^{\varepsilon t}(1-q)^{2t}}$, lo que le permitió concluir que la segunda distribución era asimilable a la primera.

La segunda distribución la obtuvo de sumar dos variables aleatorias que representan errores independientes, cada uno con distribución de probabilidades de la forma:

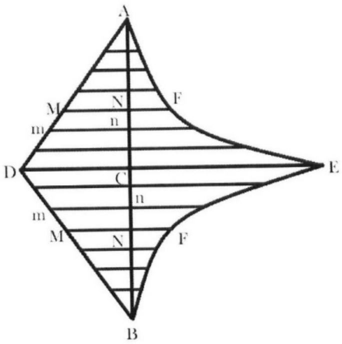
$$q^{-\varepsilon/2}, q^{-(\varepsilon-1)/2}, \dots, q^{0}, \dots, q^{(\varepsilon-1)/2}, q^{\varepsilon/2}$$

Simpson era consciente (según lo expresó en "*On the advantage of taking the mean of a number of observations in practical astronomy"* (1755)) del carácter restrictivo de las hipótesis que había asumido acerca de la distribución de los errores, pero también dejó planteada su convicción sobre la posibilidad de generalización de los resultados obtenidos. En 1757, a partir de algunas observaciones realizadas por Bayes a dicha publicación, estableció dos principios que caracterizaron los alcances de dicha generalización y que se constituyeron, a partir de ese momento, en las hipótesis implícitas en todos los trabajos sobre teoría del error: **i)** que las probabilidades de ocurrencia de errores de la misma magnitud, por defecto o por exceso, eran iguales; **ii)** que era posible definir límites determinados entre los cuales se podía suponer que estarían incluidos los errores[9].

A partir de este último postulado y remitiéndose a la primitiva distribución de probabilidades discreta proporcional a:

$$1,2, \dots, (\varepsilon+1), \dots, 1$$

introdujo una distribución triangular de los errores (ABD) (posiblemente, la primera publicación de una distribución de probabilidades continua para los errores).




La curva AFEFB representa la función de probabilidades del error que se comete al utilizar el valor medio de los errores. Integrando esta función Simpson obtuvo finalmente que la probabilidad de que el valor medio de $n$ errores no excediera una proporción dada, $1-\frac{x}{n}$, del error máximo posible (CA) estaba dada por:

$$1-\frac{2}{(2n)!}\left[x^{2n}-2n(x-1)^{2n}+\frac{2n}{1}\frac{(2n-1)}{2}(x-2)^{2n}-\right.$$

$$\left.-\frac{2n}{1}\frac{(2n-1)}{2}\frac{(2n-2)}{3}(x-3)^{2n}+\cdots\right]$$

Una circunstancia que colaboró para que Simpson alcanzara el suceso que no lograron ni los Bernoulli ni de Moivre y planteara los fundamentos de una teoría de los modelos estocásticos fue, sin duda, que se independizara de los tradicionales esquemas de urnas

[9] Simpson (1757) especificó -arbitrariamente- límites para su distribución de los errores "*...que dependen de las propiedades del instrumento y de la habilidad del observador"* (p. 64).

(en los que, si bien los argumentos matemáticos necesarios eran simples, los aspectos conceptuales inherentes al problema de la inversión de la probabilidad, dadas su estructura física intrínsecamente discreta y la asimetría existente entre la composición de la urna y el mecanismo de extracción, eran muy complicados) y seleccionara otras aplicaciones -por ejemplo, las que se derivaban de las observaciones astronómicas- en las cuales la cuestión conceptual de la inversión de la probabilidad era más fácil de aprehender.

### 3.- Laplace y la continuación de la obra de Simpson

Laplace aplicó los resultados obtenidos en la inversión de la probabilidad al problema de la selección del mejor valor promedio para un conjunto de observaciones, dada una distribución "a posteriori" de los errores ($p\left(Y\ /\ \hat{Y}\right)$). En principio Laplace propuso dos opciones: **i)** adoptar el valor de $\varepsilon = Y - \hat{Y} = \varepsilon_{me}$ (la mediana de la distribución "a posteriori" de los errores, "*media de la probabilidad*" en la nomenclatura de Laplace) tal que la probabilidad de que el verdadero valor de $\varepsilon$ esté por encima o por debajo de él, sea la misma, **ii)** adoptar el valor de $\varepsilon$ que minimice la suma de los valores absolutos de los desvíos respecto a un valor $\varepsilon_0$, multiplicados por sus correspondientes probabilidades, es decir, el adoptar el valor $\varepsilon_0$ tal que:

$$\int_{-\infty}^{\infty} |\varepsilon - \varepsilon_0| f(\varepsilon) d\varepsilon = mínimo$$

y luego demostró que ambos valores son iguales ( $\varepsilon_{me} = \varepsilon_0$).

Formalizó, asimismo, el principio Simpsoniano que establece que la cuestión central del problema de la selección del valor medio radica en la definición "a priori" de la distribución de probabilidades de los errores: Sea $Y$ el verdadero valor de la medida que se está calculando y sean $\{\hat{Y}_i\}$ $(i = 1,2,\dots,n)$ las observaciones a realizar sobre dicho valor. Si se supone que los errores $\varepsilon_i = Y - \hat{Y}_i$ poseen un comportamiento aleatorio de acuerdo con una distribución de probabilidades dada, $f(\varepsilon_i)$, entonces la probabilidad de que las observaciones asuman los valores $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n$ estará dada por:

$$p\left[\left(\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n\right) / Y\right] = f\left(Y - \hat{Y}_1\right) f\left(Y - \hat{Y}_1\right) f\left(Y - \hat{Y}_2\right) \dots f\left(Y - \hat{Y}_n\right) = f(\varepsilon_1) f(\varepsilon_2) \dots f(\varepsilon_n)$$

y la distribución de probabilidades inversa de $Y$ condicionada por $\hat{Y}$ ($p\left(Y\ /\ \hat{Y}\right)$, de acuerdo con lo demostrado en el teorema de Bayes, debe cumplir la condición de proporcionalidad, $p\left(Y\ /\ \hat{Y}\right) \propto p\left(\hat{Y}\ /\ Y\right)$ o, más estrictamente, suponiendo una distribución "a priori" de $Y$ uniforme, debe ser:

$$p\left[Y\ /\ \left(\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n\right)\right] = \frac{f(\varepsilon_1) f(\varepsilon_2) \dots f(\varepsilon_n)}{\int f\left(\hat{Y}_1 - y\right) f\left(\hat{Y}_2 - y\right) \dots f\left(\hat{Y}_n - y\right) dy}$$

Este problema de la selección de la distribución de los errores fue tratado por Laplace fundamentalmente en dos trabajos: "*Mémoire sur le probabilité des causes par les èvénements*" (1774) y "*Mémoire sur le probabilités*" (1781). Laplace comenzó aplicando el principio de la razón insuficiente a las probabilidades correspondientes a todos los valores de $\varepsilon$, con lo que obtuvo como solución la curva $f(\varepsilon) = constante$ y la desestimó porque contradecía el criterio Galileano-Simpsoniano-Newtoniano[10] de que es menos probable

[10] Debe tenerse en cuenta que, a partir de 1755 todo análisis dirigido a la determinación "a priori" de la forma de la curva representativa de la probabilidad de los errores está "obligado" a cumplir las siguientes condiciones insoslayables en un esquema post-Galileano, post-Simpsoniano y post-Newtoniano: **i)** que la curva sea simétrica con respecto a cero; **ii)** que la curva decrezca hacia cero a medida que el valor del error se aparte de cero en

cometer errores "grandes" que errores "pequeños". Luego, aplicando el principio de la razón insuficiente al decrecimiento de $f(\varepsilon)$, $\frac{d}{d\varepsilon}f(\varepsilon) = constante$, obtuvo la distribución triangular y también la desestimó porque "*Podemos ver fácilmente que esta disminución no puede ser constante; debe ser menor a medida que las observaciones se aparten más del verdadero valor*" (1774, p. 650).

En su tercera propuesta Laplace llevó el principio de la razón insuficiente mucho más allá que sus predecesores al postular que no existía ninguna razón para suponer que las ordenadas de la curva debían decrecer en forma distinta a sus decrementos (aplicación del principio de la razón insuficiente a las tasas de variación de la función):

$$\frac{df(\varepsilon + d\varepsilon)}{df(\varepsilon)} = \frac{f(\varepsilon + d\varepsilon)}{f(\varepsilon)}$$

Es decir, $\frac{df(\varepsilon)}{d\varepsilon} = -af(\varepsilon)$. El resultado obtenido fue la función doble exponencial $f(\varepsilon) = \frac{a}{2}e^{-a|\varepsilon|}$ $(-\infty \leq \varepsilon \leq \infty)$.

Esta solución significó un avance muy importante en la medida que constituyó el primer tratamiento analítico de los errores de observación. Contrariamente a lo realizado por Simpson y Legendre –quienes desarrollaron un procedimiento para la selección y evaluación del valor medio "a priori" de las observaciones-, Laplace consideró la selección del valor medio, condicionado por el conjunto de observaciones.

En la "*Mémoire sur les probabilités*"(1781), Laplace, basándose en una partición aleatoria del intervalo unitario (método que garantiza el cumplimiento del principio de la razón insuficiente, en la medida que no privilegia "a priori" a ninguna distribución de probabilidades) y preservando el cumplimiento de las hipótesis de simetría y unimodalidad, obtuvo la función:

$$f(\varepsilon) = \frac{1}{2a} log\left(\frac{a}{|\varepsilon|}\right) \qquad\qquad (|\varepsilon| \leq a)$$

(donde $a$ denota el límite de los posibles errores)[11]. Desafortunadamente esta solución también resultó poco apropiada en las aplicaciones[12].

D. Bernoulli ("*Diiudicadtio maxime probabilis plurium observationem discrepatium atqu verosimillina inductio inde formanda*" (1778)) –influido por el principio Simpsoniano de que la curva representativa de la distribución de los errores debe ser de rango finito y debe decrecer "abruptamente" hacia sus extremos- propuso un semicírculo de diámetro conocido cuyo centro era el punto estimado mediante el método de máxima verosimilitud y, en "*Specimen philosophicum de compensationibus horologicis, et veri mensura temporis*" (1780) utilizó la distribución de probabilidades binomial-simétrica y su aproximación Normal para describir los errores aleatorios en la medición de los tiempos en un reloj de péndulo. Esta fue, probablemente, la primera aplicación de la función Normal como representación de la

---

cualquiera de los dos sentidos; **iii)** que el área comprendida bajo la curva sea igual a uno; iv) que se cumpla el principio de la razón insuficiente, es decir, que la curva sea en algún sentido reducible a una descripción en términos de casos igualmente probables.

[11] La explicación de Laplace acerca de las razones para la selección de esta curva puede ser resumida en el siguiente pasaje de la "*Mémoire*": "*Es natural pensar que los mismos errores, en más y en menos, son igualmente probables y que su facilidad de ocurrencia es tanto menor cuanto más grandes sean; si no se tiene otra información con relación a su facilidad de ocurrencia se cae evidentemente en el caso precedente: es necesario, por lo tanto, suponer que en tal caso la probabilidad, tanto del error positivo $x$, como del error negativo $-x$, es igual a $(1 / 2a)log(a / x)$ y que ésta es la ley de probabilidades que hay que tomar como punto de partida para la selección del promedio de los resultados de un conjunto de observaciones*" (p. 166).

[12] Dale (1991): "*Laplace expresa que su objeto ha sido más bien hacer conocer que laluz de la teoría de la probabilidad puede iluminar este tipo de cuestiones, que presentar a los observadores un método práctico y de fácil utilización*" (p. 152).

distribución de los errores, basándose en la hipótesis que cada error puede ser considerado como la agregación de un gran número de "*errores elementales*".

## 4.- El método de la combinación de observaciones

Como una continuación de los trabajos de Simpson y asociada a cuestiones generadas en el ámbito de la astronomía y la geodesia, surgió el método conocido como de la combinación de observaciones, que tuvo su culminación con la aparición, en 1805, del principio de los cuadrados mínimos de Legendre el cual, dotado por Laplace y Gauss de connotaciones probabilísticas proporcionadas por la teoría de los errores de observación, originó la que se dio en llamar la "síntesis de Gauss-Laplace".

En 1722, R. Côtes en su trabajo sobre la estimación de los errores en mediciones trigonométricas ("*Æstimatio errorum in mixta mathesi, per variationes partium trianguli plani et spherici*") –sin mencionar las razones que lo condujeron a tal conclusión- desarrolló la siguiente solución para la determinación de la ubicación de un punto, $x$, en un plano: Dadas cuatro observaciones, $x_1, x_2, x_3$ y $x_4$, realizadas sobre dicho punto y, dadas las respectivas ponderaciones, $d_1, d_2, d_3$ y $d_4$ (inversamente proporcionales a las distancias entre las $x_i$ y $x$), "*la ubicación más probable del punto $x$ es la proporcionada por el promedio aritmético ponderado de las observaciones*" (p. 107)[13].

De acuerdo con Laplace, la primera aplicación que se conoce del resultado de Côtes se debe a Euler (1749) al introducir el hoy conocido como "método de los promedios", tomando como punto de partida una ecuación de la forma[14]:

$$\varepsilon = Y - a_1X_1 - a_2X_2 - \cdots - a_6X_6$$

Donde $\varepsilon$ denotaba el "error de observación aleatorio", los elementos $Y$ y $X_j$ $(j = 1,2,\dots,6)$ eran obtenidos a partir de la observación y los coeficientes $a_j$ $(j = 1,2,\dots,6)$ constituían las incógnitas. A partir de las observaciones obtuvo 75 ecuaciones para estimar 6 incógnitas. Intentó, entonces, resolver el problema de la sobreestimación del sistema formando tantos conjuntos de ecuaciones como incógnitas tenía que resolver. Estos conjuntos estaban estructurados a partir de una adecuada selección en pequeños conjuntos de ecuaciones tomadas bajo condiciones similares de modo que proporcionaran, esencialmente, coeficientes aproximadamente iguales y tales que la diferencia entre dos ecuaciones cualesquiera produjera la eliminación de muchos términos. A partir de las 6 ecuaciones resultantes, estimó los coeficientes $a_j$ como aquellos valores que minimizaban el mayor de los errores.

Al año siguiente de la publicación de Euler, apareció "*Abhandlung über die Umwalzung del monds um seine Axe ind die Scheinbare Bewengung der Mondsflecten*" de T. Mayer (1750) quien, a partir de la expresión:

$$\varepsilon = Y - a_0 - a_1X_1 - a_2X_2$$

y de las observaciones con que contaba para los elementos $Y, X_1$ y $X_2$, generó un sistema de 27 ecuaciones lineales en las incógnitas $a_0, a_1$ y $a_2$, a las que luego agrupó, basándose en los valores que asumían los coeficientes de las incógnitas, en tres conjuntos de 9 ecuaciones cada uno. Sumando las ecuaciones incluidas en cada grupo, definió 3 ecuaciones lineales para las 3 incógnitas, cuyas estimaciones fueron obtenidas minimizando el error máximo.

---

[13] La referencia bibliográfica corresponde a la traducción de Gowing (1983).

[14] El método natural de estimación consistía en construir las denominadas "*ecuaciones de condición*", igualando los valores estimados y observados y resolviendo los valores de los coeficientes a partir de un sistema con tantas ecuaciones como coeficientes a estimar.

Una vez resuelto el problema del agrupamiento de las ecuaciones, Mayer planteó el problema de la medida de la confiabilidad de las estimaciones obtenidas. Para ello formuló una teoría de los errores que, utilizando notación moderna, podría ser resumida de la siguiente forma: Sea $\varepsilon$ el error que se comete al tomar como valor de un coeficiente $a$ a la estimación $\hat{a}_1$, obtenida como solución de un sistema de $n_1$ ecuaciones y sea $\hat{a}_2$ la estimación del mismo coeficiente obtenida de un sistema de ecuaciones de tamaño $n_2$. Partiendo de la hipótesis de que el error en cada una de las estimaciones varían en forma inversamente proporcional al número de observaciones[15], concluyó que:

$$\frac{\varepsilon}{1 / n_1} = \frac{|\hat{a}_2 - \hat{a}_1| + \varepsilon}{1 / n_2}$$

Es decir, que $\varepsilon = \frac{n_2}{n_1 - n_2} |\hat{a}_2 - \hat{a}_1|$.

Como se puede apreciar, la aproximación estadística de Mayer a la solución del problema de las observaciones realizadas en condiciones diferentes, es totalmente diferente a la solución matemática de Euler. La hipótesis de Mayer de que la combinación de observaciones aumentaba la confiabilidad de los resultados proporcionalmente al número de ecuaciones combinadas –contraria al criterio de Euler, que considera que el error aumenta con la agregación-, se basaba en el concepto estadístico que supone la compensación de los errores aleatorios, de acuerdo con los principios Simpsonianos-Newtonianos.

En 1755 R.J. Boscovich, planteó una solución basada en cinco ecuaciones del tipo:

$$Y = a_0 + a_1 X_1 + \varepsilon$$

El gran avance de Boscovich con respecto al trabajo de Mayer fue la formulación de un principio objetivo general de acuerdo con el cual los coeficientes $a_0$ y $a_1$ debían ser estimados de acuerdo con las siguientes condiciones:

**i)** $\sum_i \varepsilon_i = \sum_i (Y_i - a_0 - a_1 X_i) = \frac{1}{n} \sum_i Y_i - a_1 \sum_i X_i = 0$

**ii)** $\sum_i |\varepsilon_i| = \sum_i |Y_i - a_0 - a_1 X_i| = min$

A partir de la primera de estas condiciones que, como se vio, después de los trabajos de Simpson y Newton, se convirtió en obligada en todo análisis “a priori” de la distribución de probabilidades de los errores y que es la consecuencia formal de la noción intuitiva de que los errores positivos y negativos deben ser igualmente probables, se obtiene que:

$$a_0 = \frac{1}{n} \sum_i Y_i - a_1 \frac{1}{n} \sum_i X_i$$

Con respecto a la segunda condición, a partir de la cual, reemplazando el valor de $a_0$ se obtiene que:

$$\sum_i \left| \left( Y_i - \frac{1}{n} \sum_i Y_i \right) - a_1 \left( X_i - \frac{1}{n} \sum_i X_i \right) \right| = min$$

[15] De acuerdo a lo demostrado por de Moivre, esta hipótesis es errónea: la confiabilidad de una estimación varía en forma directamente proporcional a la raíz cuadrada del número de observaciones, suponiendo que las observaciones se hayan realizado en igualdad de condiciones. No obstante es necesario destacar la originalidad de la idea que precedió en mucho tiempo a los desarrollos de Laplace y Gauss.

Boscovich interpretó que la única solución analítica posible era mediante la diferenciación con respecto a $a_0$ y $a_1$, por lo que eligió desarrollar una solución geométrica. Fue Laplace (1786) quien estableció un tratamiento algebraico del algoritmo de Boscovich (al cual denominó "*método de situación*".

Como continuador inmediato de la línea teórica de Simpson merece ser mencionado J.H. Lambert (1760)(1765). En particular en "*Beyträge zum Gebrauche der Mathematik und deren Anwendung*" (1765), enunció condiciones específicas sobre las distribuciones de probabilidades de los errores y desarrolló métodos de estimación que, si bien no fueron ampliamente adoptados, ejercieron cierta influencia sobre la obra de Laplace y Gauss.

## 5.- La síntesis de Gauss-Laplace y el criterio de optimización de los cuadrados mínimos

Las dos corrientes teóricas expuestas en la sección precedente (Mayer-Laplace y Boscovich-Laplace) –originadas, según se vio, en la premisa de que la combinación de observaciones era esencial para la interpretación de la realidad a partir de la experiencia- tuvieron su culminación en la obra del matemático francés A.M. Legendre. En 1805 este autor dio a conocer "*Nouvelles méthodes pour la détermination des orbites des comètes*", en cuyo apéndice –titulado "*Sur la méthode des moindres carrés*"- expuso el método de los cuadrados mínimos, según el cual "*En muchas investigaciones, en la que el objetivo es deducir resultados a partir de observaciones en la forma más confiable posible, nos encontramos con un sistema de ecuaciones de la forma:*

$$y_i = ax_i + bz_i + cw_i + \cdots + \varepsilon_i$$

*donde $y_i, x_i, z_i, w_i, \dots$ denotan coeficientes de valor conocido, que varían de una ecuación a otra y $a, b, c, \dots$ denotan cantidades desconocidas a ser determinadas de acuerdo con la condición de que todos los valores de $\varepsilon_i$ (que representan a los errores de obervación) se anulen, o se reduzcan a una cantidad muy pequeña*" (p. 77).

La solución que propuso consistía " *en minimizar la suma de los cuadrados de los errores*". Legendre obtuvo esta minimización inmediatamente, resolviendo el sistema de ecuaciones lineales que resultaba de derivar dicha suma de los cuadrados de los errores:

$$f(a, b, c, \dots) = \sum_i \varepsilon_i^2 = \sum_i (y_i - ax_i - bz_i - cw_i - \cdots)^2$$

con respecto a $a, b, c, \dots$[16] :

$$\sum_i x_i y_i = a \sum_i x_i^2 + b \sum_i x_i z_i + c \sum_i x_i w_i + \cdots$$

$$\sum_i z_i y_i = a \sum_i z_i x_i + b \sum_i z_i^2 + c \sum_i z_i w_i + \cdots$$

$$\sum_i w_i y_i = a \sum_i x_i w_i + b \sum_i z_i w_i + c \sum_i w_i^2 + \cdots$$

[16] En realidad, Legendre no proporciona ningún tipo de demostración teórica. Sólo comenta que " *para definir la ecuación del mínimo con respecto a cada una de las incógnitas, es necesario multiplicar todos los términos de cada ecuación por el coeficiente de dicha incógnita en esa ecuación tomado con su propio signo, y luego hallar la suma de todos estos productos*" (p. 77).

Según se vio en las secciones precedentes, durante el siglo XVIII se habían desarrollado, en forma completamente independiente, dos metodologías dirigidas a resolver el problema de la inferencia. Por una lado, una corriente que evolucionó a partir del análisis de la distribución de probabilidades binomial y que culminó con los trabajos de Laplace referidos al establecimiento de los principios para la selección del valor medio de un conjunto de observaciones y de las distribuciones de probabilidades representativas de los errores de observación. Y, por otro, una línea basada en los trabajos de Mayer-Laplace y Boscovich-Laplace, sobre combinación de observaciones, que derivó en el método de los cuadrados mínimos de Legendre y que se desarrolló atendiendo exclusivamente a los aspectos matemáticos del problema, sin ninguna consideración formal sobre sus aspectos probabilísticos, ni sobre las cuestiones referidas a la cuantificación de la incertidumbre, que son inherentes a la inferencia[17].

Fue Gauss el encargado de conectar ambas corrientes teóricas. En 1809 se publicó su "*Theoria motus corporum coelestiorum in sectionibus conicis solum ambientium*", la cual contiene una sección dedicada a su interpretación del problema de la combinación de observaciones, en la que presenta la siguiente exposición del método de los cuadrados mínimos: Sean $n$ funciones lineales de la forma:

$$\begin{cases} \hat{Y}_1 = a_1 X_{11} + a_2 X_{21} + a_3 X_{31} + \cdots \\ \hat{Y}_2 = a_1 X_{12} + a_2 X_{22} + a_3 X_{32} + \cdots \\ \hat{Y}_3 = a_1 X_{13} + a_2 X_{23} + a_3 X_{33} + \cdots \\ \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \end{cases}$$

(donde $X_1, X_2, X_3, \ldots$ son observables y $a_1, a_2, a_3, \ldots$ representan $n$ coeficientes desconocidos cuyos valores pueden ser obtenidos a partir de un conjunto de observaciones $Y_1, Y_2, Y_3, \ldots$; $X_{11}, X_{12}, X_{13}, \ldots$; $X_{21}, X_{22}, X_{23}, \ldots$) que integran el sistema de "*ecuaciones normales*" (de acuerdo con la nomenclatura de Gauss (1823))[18].

Si se supone que los errores $\varepsilon_i = Y_i - \hat{Y}_i \ (i = 1,2,3, \ldots)$ son variables iid con función de densidad $(f(\varepsilon_i) = f(\varepsilon) \ (i = 1,2,3, \ldots)$, entonces el sistema de valores "más probable" corresponderá al conjunto de coeficientes $a_1, a_2, a_3, \ldots$ que maximiza, de acuerdo con el método de optimización de máxima verosimilitud, la función $\varphi = f(\varepsilon_1) f(\varepsilon_2) f(\varepsilon_3) \ldots$. Estos valores pueden ser obtenidos a partir de la resolución del sistema de ecuaciones $\frac{\partial \ln \varphi}{\partial a_i} = 0$ $(i = 1,2,3, \ldots)$. Ahora bien, esta solución implica necesariamente la definición previa de la función $f(\varepsilon)$. Pero la forma de esta distribución no puede ser determinada "a priori", sólo se le pueden atribuir las ya mencionadas características Galileanas-Simpsonianas-Newtonianas de carácter general. A partir de estas hipótesis y asumiendo como axioma el principio de que la media aritmética de un conjunto de observaciones realizadas en igualdad de condiciones sobre una medida, constituye el valor más probable de la misma, se demuestra que, en el caso en que se verifique que $Y_i = a + \varepsilon_i$, entonces la media aritmética de las observaciones:

$$\bar{Y} = \frac{1}{n}(Y_1 + Y_2 + Y_3 + \cdots)$$

maximiza la función $\varphi$ si y sólo si:

---

[17] Stigler, S.M. (1975): "El método de los cuadrados mínimos proporcionó resultados que pudieron ser considerados 'los mejores', ya que minimizaban la suma de los cuadrados de los errores y producían un atractivo equilibrio mecánico; pero, dado que la naturaleza estocástica de las observaciones no fue tenida en cuenta, es decir, no se consideró la cuantificación de la incertidumbre, no fue posible dar una respuesta a la pregunta ¿Cuán bueno es 'el mejor´?"

[18] Gauss no proporciona ninguna explicación acerca de los detalles de esta demostración.

$$f(\varepsilon) = \frac{h}{\sqrt{\pi}} e^{-h^2\varepsilon^2}$$

donde $h > 0$ denota la medida de la precisión de las observaciones.

A partir de la versión de Laplace del teorema de Bayes (suponiendo, de acuerdo con el principio de la razón insuficiente, una distribución uniforme de las probabilidades “a priori”), Gauss demostró que la distribución “a posteriori” de las incógnitas cumple la siguiente condición de proporcionalidad:

$$p\{[(a_1 < \hat{a}_1 < a_1 + da_1) \cap (a_1 < \hat{a}_1 < a_1 + da_1) \cap \ldots] \,/\, [(Y_1 = y_1) \cap (Y_2 = y_2) \cap \ldots]\} =$$

$$= k\varphi da_1 da_2 \ldots = \frac{\varphi da_1 da_2 \ldots}{\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \varphi da_1 da_2 \ldots} \approx e^{h^2 \sum \varepsilon^2}$$

y que esta probabilidad se maximiza cuando se minimiza la función $\sum_i \varepsilon_i^2$. Es decir, cuando se verifica el principio de los cuadrados mínimos[19].

En su “*Bestimmung der Genauigkeit der Beobachtungen*” (1816) Gauss introdujo la función:

$$F(u) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^{u} e^{-u^2} du$$

y el coeficiente $r = \frac{\theta}{h}$ (donde $\theta = 0{,}4769363$ es el valor tal que $F(\theta) = \frac{1}{2}$) representa el “*error probable para la función* $F(hu)$”.

Analizó, por otra parte, la estimación de $h$ a partir del resultado de $n$ observaciones independientes. Diferenciando la función de densidad correspondiente al sistema de las $n$ observaciones:

$$f(\varepsilon) = ch^n exp\left(-h^2 \sum_i \varepsilon_i^2\right)$$

obtuvo que:

$$f'^{(\varepsilon)} = ch^{n-1}\left(n - 2h^2 \sum_{i=1}^{n} \varepsilon_i^2\right) exp\left(-h^2 \sum_i \varepsilon_i^2\right)$$

y concluyó que el valor de $h$ correspondiente al máximo de la función $f(\varepsilon)$ estaba definido por:

[19] Desde el primer momento se planteó una ardua disputa entre Gauss y Legendre por la paternidad del método de los cuadrados mínimos. De acuerdo con su propio testimonio Gauss (según la correspondencia de Gauss a H.M.W. Olbers, a B.A. von Lindenau y a F.X. von Zach de 1802) había utilizado esta metodología desde 1795. Pero, aparentemente, no adivinó (o no logró comunicar) la importancia del resultado. Su conclusión fue que no era posible lograr ningún progreso en términos de estimaciones “*más probables*” de las incógnitas, si no se definía la forma de la distribución de probabilidades de los errores, por lo que decidió adoptar el camino inverso y buscar la distribución que convirtiera a la media artmética en el valor más probable. Ahora bien, más allá de la cuestión de la prioridad, es necesario tener en cuenta que la contribución de Legendre se limitó a la definición del principio de los cuadrados mínimos y que fue Gauss el encargado de desarrollar la teoría basada en dicho principio.

$$h=\sqrt{\frac{n}{2\sum_{i=1}^{n}\varepsilon_i^2}}$$

Este resultado le permitió demostrar que el valor más probable de $r$ era:

$$r=\frac{\theta}{h}=\theta\sqrt{\frac{2\sum_i\varepsilon_i^2}{n}}$$

Posteriormente consideró el caso más general para varianzas no-constantes ($\sigma^2(\varepsilon_i)$, $i=1,2,\ldots$), lo que lo condujo al método de los "cuadrados mínimos ponderados", en el que la función a minimizar era de la forma $\sum_i\frac{\varepsilon_i^2}{\sigma_i^2}$ y amplió los desarrollos anteriores a funciones lineales que involucraban más de una incógnita, demostrando que el vector de los coeficientes $a=[a_i]$ poseía una distribución de probabilidades Normal multidimensional del tipo $N(a,(XX')\sigma_\varepsilon^2)$.

El razonamiento de Gauss fue exactamente el inverso del propuesto en su momento por Simpson y Laplace quienes, basándose en el principio de la razón insuficiente, establecieron condiciones para determinar la forma "a priori" de la curva $f(\varepsilon)$ y, a partir de esta definición, seleccionar promedio más adecuado para utilizar como verdadero valor del error. Por el contrario, el argumento de Gauss pude ser resumido de la siguiente forma: **i)** la media aritmética de un conjunto de observaciones referidas a una medida desconocida constituye el valor más probable de dicha medida, sólo si los errores se distribuyen de acuerdo a una función Normal; **ii)** por su parte, la media aritmética es generalmente reconocida como la mejor forma de "*combinar*" observaciones, por lo que se puede concluir que los errores deben ser interpretados como Normalmente distribuidos; iii) pero la hipótesis de Normalidad de los errores permite la maximización de la función $\varphi$, cuando se minimiza la forma cuadrática $\sum\varepsilon_i^2$ y esto se verifica cuando los valores de $a_i$ son obtenidos de acuerdo con el principio de los cuadrados mínimos[20].

Aquí hizo su aparición una vez más Laplace para enmendar la falla epistemológica que implica la aparente circularidad de este razonamiento. En el suplemento a su "*Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités*" (1786), partir del supuesto que los errores a los que se refería la formulación de Gauss estaban formados por la agregación de "*errores elementales*" y, vinculando su versión del teorema central del límite con la estimación lineal, demostró que era legítimo suponer que dichos errores poseyeran una distribución asintóticamente Normal. Por otra parte, basándose en su "*Mémoire sur la probabilité des causes par les èvénements*" (1774), concluyó que las estimaciones "*más probables*" obtenidas por el método de los cuadrados mínimos minimizaban, a su vez, el valor esperado "a posteriori" del error.

En 1811 Laplace publicó la "*Mémoire sur les intégrales définies et leur application aux probabilités, et spécialement á la recherche du milieu qu'il faut choisir entre les résultats des observations*", que contiene una formulación algebraica de su generalización del método de Mayer: Sea un sistema de ecuaciones inconsistentes de la forma $\varepsilon_i=aX_i-Y_i$ $(i=1,2,3,\ldots)$. Donde los valores de $X_i$ e $Y_i$ se obtienen de la observación repetida en igualdad de condiciones, $a$ es una incógnita y los $\varepsilon_i$ denotan variables aleatorias que representan a los errores no-sistemáticos y simétricamente distribuidas con respecto a cero. Laplace construyó tantas combinaciones lineales de las ecuaciones como incógnitas a resolver:

[20] Una derivación alternativa –prácticamente desconocida-de la función Normal como "*ley para la distribución de los errores aleatorios*" fue propuesta por Adrain (1809).

$$\sum_i k_i\varepsilon_i = a\sum_i k_iX_i - \sum_i k_iY_i$$

De modo que:

$$a = \frac{\sum_i k_iY_i}{\sum_i k_iX_i} + \frac{\sum_i k_i\varepsilon_i}{\sum_i k_iX_i} = \frac{\sum_i k_iY_i}{\sum_i k_iX_i} + u$$

Donde $u$ denotaba la variable aleatoria que representaba el error que surgía cuando la función lineal de los errores ($\sum k_i\varepsilon_i$) no era exactamente igual a cero.

Si el número de observaciones era suficientemente grande, de acuerdo con una presentación ligeramente diferente del teorema central del límite, esta variable tendría una distribución aproximadamente Normal con valor esperado proporcional a $f(k_i) = \frac{\sqrt{\sum k_i^2}}{\sum k_iX_i}$. A partir de la relación:

$$\frac{df}{dk_i} = \frac{\sum_i k_i \sum_i k_iX_i - \sum_i k_i^2 \sum_i X_i}{\sqrt{\sum_i k_i^2}\,(\sum_i k_iX_i)^2} = 0$$

Laplace concluyó que el "*error esperado*" asumía su valor mínimo cuando las ponderaciones $k_i$ eran proporcionales a $X_i$ y, en ese caso se verificaba que $a = \frac{\sum_i X_iY_i}{\sum_i X_i^2}$ la cual era, precisamente, la estimación que se obtenía de aplicar el método de los cuadrados mínimos; es decir, de minimizar la forma cuadrática $(a\sum_i X_i - \sum_i Y_i)^2$ con respecto a $a$. En otras palabras, Laplace demostró: **i)** que todos los estimadores de $a$ definidos como funciones lineales de los valores observados $Y_i$ poseían una distribución de probabilidades aproximadamente Normal y **ii)** que entre esos estimadores, los que se obtenían de aplicar el método de los cuadrados mínimos, eran los que poseían menor error esperado.

En esa misma memoria Laplace generalizó este resultado a casos que incluían más de una incógnita:

$$\varepsilon_i = a_1X_{i1} + a_2X_{i2} - Y_i \qquad\qquad (i = 1,2,3,\dots)$$

y demostró que las estimaciones que minimizaban el error esperado eran las que proporcionaba el método de los cuadrados mínimos, $\hat{a} = (X^TX)^{-1}X^TY$ (donde $X = [X_{ij}]$ e $Y = [Y_i]$). Es decir las que minimizaban la expresión $\sum_i(a_1X_{i1} + a_2X_{i2} - Y_i)^2$.

Por su parte, en 1823, Gauss publicó la primera parte de su "*Theoria combinationis observationum erroribul minimis obnoxiæ*" en la que, abandonando su razonamiento circular de 1809, expuso una generalización del método de Laplace. Demostró que entre todas las estimaciones insesgadas de los coeficientes, aquellas obtenidas de acuerdo con el criterio de los cuadrados mínimos poseían varianza mínima. En la segunda parte de esta obra – publicada en 1823- generalizó los resultados anteriores a cualquier combinación lineal de los coeficientes. Halló la fórmula para determinar la suma de los cuadrados de los residuos ($\hat{\sigma}_\varepsilon^2$) a partir de las estimaciones de los $a$. Desarrolló un procedimiento para permitir el agregado al modelo de un $a$ adicional sin tener que recalcular los coeficientes ya estimados. Demostró que, para series de $n$ observaciones en modelos con $k$ incógnitas ($n > k$), se verificaba que $E(\hat{\sigma}_\varepsilon^2) = (n-k)\sigma_\varepsilon^2$.

El núcleo central alrededor del cual se desarrolló esta segunda versión de la teoría de los cuadrados mínimos podía resumirse en el siguiente principio: Cuando se utiliza un estimador $\hat{a}$ de $a$, se comete un error $a-\hat{a}$ que implica una pérdida. Se debe elegir, entonces un estimador que minimice dicha pérdida. Si se supone que la pérdida es proporcional al error cuadrático $(a-\hat{a})^2$, el estimador debe ser el que minimice el error medio cuadrático $E[(a-\hat{a})^2]$.

Gauss demostró que si se acepta la hipótesis de que los errores: i) son lo suficientemente pequeños como para que sus potencias de orden mayor que uno pueden ser despreciadas (es decir, si se consideran solamente los estimadores lineales, ii) están independientemente distribuidos con valor esperado nulo ($E(\varepsilon)=0$) y la misma varianza desconocida ($E(\varepsilon\varepsilon^T)=\sigma_\varepsilon^2 I$), entonces los estimadores obtenidos a partir de la aplicación del principio de los cuadrados mínimos, son los que minimizan el error medio cuadrático $(X^TX)^{-1}\sigma_\varepsilon^2$ .

## 6.- Modelos econométricos y procesos estocásticos

A partir de 1821 la síntesis de Gauss-Laplace había alcanzado el grado de madurez necesario como para afirmar que el problema de la inferencia se encontraba resuelto y en condiciones de ser aplicado a ramas de la ciencia muy distintas de la astronomía y la geodesia en las que se había generado. No obstante y a pesar de la innegable importancia en el análisis de los fenómenos (ampliamente reconocida desde la época de Bernoulli), la teoría de la inferencia necesitó más de un siglo y el desarrollo de los conceptos de correlación y regresión para que las aplicaciones se extendieran a las ciencias sociales, fundamentalmente por obra de J. Fourier, F. Galton, G. Weldon, F.Y. Edgeworth, K. Pearson, E.S. Pearson, J. Neyman, G.U. Yule y R.A. Fisher.

Esta demora en la adopción de una interpretación probabilística del comportamiento de los fenómenos de las ciencias sociales, parece confirmar el profundo carácter filosófico de este cambio intelectual y, fundamentalmente, el peso del determinismo en la historia de las ciencias fácticas. Resulta sumamente curioso que autores como A. Quetelet, A. Cournot y W.S. Jevons, quienes publicaron obras sobre teoría de la probabilidad, no utilizaran técnicas probabilísticas en las ciencias sociales.

Esta actitud de los humanistas de la primera mitad del siglo XIX puede ser atribuida principalmente a la tradicional identificación de los conceptos de probabilidad y frecuencia relativa y, en consecuencia, al problema de categorización de las observaciones en grupos homogéneos que permitían construir los “colectivos”. Esta categoría, que en astronomía puede suponerse “a priori”, en las aplicaciones económicas sólo puede aceptarse eventualmente “a posteriori” con un grado de complicación, obviamente, mucho mayor.

Como consecuencia de esta modificación en los fundamentos de la inferencia, se produjo el abandono del teorema de Bayes, no sólo como un rechazo a nivel teórico, sino también debido a que esta nueva forma de evaluación de la probabilidad no permitía, salvo en casos muy particulares, la determinación de las probabilidades “a priori”. Esta circunstancia se vio agravada por la negación del principio de indiferencia y, en consecuencia, de la hipótesis de distribución uniforme de las probabilidades “a priori”.

Esta modificación dogmática no implicó solamente el rechazo de la distribución uniforme sino, también, la pérdida de la hegemonía de la distribución Normal. En contradicción a las propuestas de Quetelet (1848) y Galton (1889) –a quienes se puede considerar los artífices del redescubrimiento de la distribución Normal en el comportamiento de gran número de fenómenos naturales- K. Pearson (1893)(1894)(1895)(1896)(1978) concluyó que la adaptación de la curva Normal a los resultados experimentales era mucho menor de lo que se había supuesto. Los resultados experimentales revelaron a Pearson (1895) que en las distribuciones empíricas era común observar leves asimetrías, circunstancia que lo condujo a definir una familia formada por 12 curvas de frecuencia que utilizó asimilándolas a distribuciones de probabilidades.

En lo que hace al ámbito particular de la teoría económica, el desarrollo de la economía matemática –iniciado a comienzos del siglo XIX- no adoptó a la probabilidad como una herramienta útil en las investigaciones hasta 1930, a pesar de que la teoría de la probabilidad –que ya había revolucionado los métodos de investigación de otras ciencias (como la biología)- se desarrolló fundamentalmente como un instrumento para analizar el comportamiento de los fenómenos de las ciencias sociales, en especial, de la economía y de la ciencia actuarial.

Entre los autores de este período de "economía sin probabilidad" merecen ser mencionados A.A. Cournot (1838) (quien puede ser considerado el primer modelista de la economía), L. Walras (1874) (quien trató a la teoría económica según los cánones de la mecánica racional, basándose en conceptos puramente determinísticos), W.S. Jevons (1865), F.Y. Edgeworth (1910), J.M. Keynes (1921).

Una explicación a esta demora en la introducción de la probabilidad en la modelización económica puede derivarse del hecho que todos estos autores desarrollaron un concepto de teoría económica basado en la interpretación determinística del comportamiento de los fenómenos de la física clásica de fines del siglo. Fue la influencia que ejercieron los trabajos de Maxwell y Gibbs (en particular sobre Edgeworth (1881), Samuelson (1947) y Bowley (1901)) y el cambio de paradigma en las investigaciones biológicas (en particular por la influencia de los trabajos de Galton y Pearson) lo que logró relacionar la teoría económica con los métodos estadísticos para cuantificar leyes y relaciones entre factores. Este avance se debió fundamentalmente a los resultados obtenidos en la estadística macroeconómica surgida de las contribuciones de De Bruyn Kops (1869), D. Baxter (1870) y F. Fellner (1901)(1905), continuadas por los por los trabajos de Marshall y Edgeworth referidos a la verificación estadística de conceptos de teoría económica y por el desarrollo de la teoría de los números índice por obra de E.A.T. Laspeyres (1864a)(1864b), H. Paasche (1871), M.W. Drobisch (1871),W.S. Jevons (1865), H.L. Wetergaard (1890), I. Fisher (1892)(1911)(1922) y F.Y. Edgeworth (1896).

Debe tenerse en cuenta, además, que esta introducción de los métodos estadísticos en la ciencia económica dio origen a la econometría de los fenómenos dinámicos la cual, basándose en el argumento que las observaciones sobre los fenómenos económicos no eran en igualdad de condiciones y, en consecuencia, no cumplían las condiciones de independencia y homogeneidad exigidas por la inferencia clásica (surgida de la síntesis de Gauss-Laplace), no sólo ignoró las nociones probabilísticas, sino que las rechazó.

Esta circunstancia condujo un método alternativo de inferencia (basado en conjuntos de observaciones obtenidas en condiciones diferentes) en el que un fenómeno dinámico $Y(t,w)$ se consdera asimilable a un proceso estocástico que evoluciona en el dominio del tiempo ( $t \in T$) y cuya configuración varía en el dominio de los estados (o de las fases o de las variables) ($w \in \Omega(Y)$).

La interpretación determinística del comportamiento de $Y(t,w)$ (al menos a nivel macroscópico) se basa en ciertas premisas de orden metafísico[21]: **i)** que el ámbito al que pertenecen los fenómenos es real; **ii)** que existen leyes objetivas que rigen su comportamiento y **iii)** que estas leyes son inherentes a los fenómenos, racionales y

[21]. A las cuales no es posible atribuir ningún fundamento (ni inductivo, ni deductivo) y que, según Daston (1988), en la interpretación de Hume (1718), constituyen "*...una necesidad psicológica un precepto casi involuntario instituido por la caritativa naturaleza para compensar las deficiencias de la razón humana*" (p. 202).

asintóticamente cognoscibles. En otros términos, cada estado $w(t)$ se supone definido por la realización simultánea, en el momento $t$, de las infinitas variables aleatorias que forman su estructura causal:

$$\Omega\big(Y(t)\big) = \{Y(t-j), X_1(t-h_1), X_2(t-h_2), X_3(t-h_3), \dots\}$$

(para $j, h_1, h_2, \dots \in \mathbb{R}, j > 0, h_i \geq 0, i = 1,2, \dots$) y la sucesión temporal de los estados determina su trayectoria. De modo que el fenómeno $Y(t, w)$ queda definido como una entidad que evoluciona en un ámbito espacio-temporal caracterizado por el principio de causalidad.

Este supuesto principio de solidaridad universal que relaciona causalmente a los fenómenos y que hace que la naturaleza de $Y(t)$ aparezca como infinitamente complicada, permite concluir que la información con que cuenta el observador ($\Omega^*\big(Y(t)\big) \subset \Omega\big(\mathrm{Y(t)}\big)$) siempre será insuficiente y que, en consecuencia, una parte importante de su comportamiento permanecerá ignorada para sí, de modo que, en ciertas condiciones de estacionariedad, se puede escribir $Y(t) = f\big[\Omega^*\big(Y(t)\big)\big] + \varepsilon(t)$, donde: **i)** $f\big[\Omega^*\big(Y(t)\big)\big]$ denota la representación del comportamiento de $Y(t)$ a partir del conjunto de información $\Omega^*\big(Y(t)\big)$, es decir, el comportamiento que debería observar $Y(t)$ si los factores incluidos en $\Omega^*\big(Y(t)\big)$ fueran sus únicas causas y no estuvieran afectados por errores de medición, y $f[\cdot]$ fuera la función que representara la verdadera relación causal invariante en el tiempo entre estos factores e $Y(t)$; **ii)** $\lim\limits_{\Omega^*(Y(t)) \to \Omega(Y(t))} f\big[\Omega^*\big(Y(t)\big)\big] = Y(t)$ y **iii)** $\boldsymbol{\varepsilon(t)}$ denota el componente azar-ignorancia ($\varepsilon(t) = Y(t) - f\big[\Omega^*\big(Y(t)\big)\big]$) y, en consecuencia, es tal que $\lim\limits_{\Omega^*(Y(t)) \to \Omega(Y(t))} \varepsilon(t) = 0$).

En este paradigma, para un observador ideal que contara con un conjunto de información no afectado por errores de medición, que abarcara la totalidad de la estructura causal de un fenómeno, su trayectoria admitiría una representación en términos de mecánica clásica reversible en el dominio del tiempo, en la que éste transcurriría uniformemente sin ninguna relación con ningún elemento externo y, en consecuencia, la diferencia entre el pasado y el futuro no poseería ningún significado, en la que, conocido el estado presente del fenómeno, se podría reproducir su pasado y calcular su futuro en forma determinística; es decir, en la que el presente no sería sino un punto que separa el pasado del futuro[22].

La insuficiencia de este modelo en términos de mecánica clásica para explicar "*...un mundo inestable que conocemos a través de una ventana finita*" (Prigogine; Nicolis (1977, p. 16), en el que el estado natural de los fenómenos es de no-equilibrio -un no-equilibrio constructivo que, como consecuencia de su propiedad fundamental de auto-organización, genera nuevos estados y nuevas estructuras complejas que sólo son imaginables en el ámbito de la irreversibilidad temporal- dio origen a una nueva formulación termodinámica -aleatorista-, cuya diferencia con la dinámica clásica radicó esencialmente: **i)** en la postulación del concepto de estado del proceso en un instante dado como resultante de una evolución orientada en el tiempo en la que, a diferencia del pasado y del presente, el futuro está formado por una sucesión de variables aleatorias no-observables vinculadas causalmente; **ii)** en la concepción de $f\big[\Omega^*\big(Y(t)\big)\big]$ como la representación de ciertas regularidades locales observadas y en la sustitución de la interpretación clásica de $\varepsilon(t)$ como azar-ignorancia (epistemológico), generado por los "errores" en la medición de las variables por la interpretación como azar-absoluto (ontológico), generado por dichos errores más las innovaciones a que están sometidos los factores incluidos en el sistema $\Omega\big(Y(t)\big)$.

---

[22]. Si se tiene en cuenta que la información infinita es indiscernible de ininteligibilidad, se puede concluir que, aún en una situación ideal en la que el observador contara con un conjunto de información de tamaño infinito, el problema de la inexplicabilidad del comportamiento de los fenómenos permanecería.

## 7.- Slutsky, Yule y la autorregresividad

A partir de esta asimilación entre fenómenos dinámicos y procesos estocásticos, la econometría evolucionó adoptando alternativamente interpretaciones determinísticas y aleatoristas de las perturbaciones. La econometría se basó inicialmente en una interpretación determinística del comportamiento de los fenómenos económicos, cuya hipótesis fundamental consistió en suponer que un modelo económico ($f[\Omega^*(Y_t)]$ $(t = 1,2,\ldots)$) constituye una representación incompleta del comportamiento de un fenómeno y que $\varepsilon_t$ representa una variable aleatoria no-observable definida por la agregación de los errores de medición que afectan a los factores incluidos en $\Omega^*(Y_t)$ y de las influencias que ejercen sobre dicho fenómeno los infinitos factores de su entorno económico no incluidos en $\Omega^*(Y_t)$[23].

Este punto de partida condujo con el tiempo a un proceso de formalización basado en el principio que la descripción de los aspectos cuantitativos de los fenómenos económicos a partir de conjuntos finitos de observaciones requería de la representación de sistemas estocásticos de variables interrelacionadas y, por lo tanto, de la teoría de la probabilidad para la estimación de dichas representaciones; es decir, para la construcción de modelos econométricos entendidos como el "*...vehículo útil para comparar la teoría económica con las observaciones, especificado de una forma medible y testeable*" (Qin (1993, p. 37) (ver Staehle (1933))[24].

Este proceso de formalización dio origen a varios problemas referidos: **i)** a la posibilidad de asegurar la adecuacidad de la representación a las relaciones teóricas; **ii)** a su identificación, es decir a la determinación de las condiciones que permitieran asegurar la unicidad de los coeficientes estimados y su interpretación, independientemente de las distribuciones de probabilidades consideradas y **iii)** a la modificación de la interpretación del término $\varepsilon_t$ asociado a la representación $f[\Omega^*(Y_t)]$ y, en consecuencia, de la componente residual ($\varepsilon_t$) asociada con el modelo correspondiente[25].

Durante el período pre-modelístico las aplicaciones econométricas consistieron en el cálculo y la predicción de relaciones económicas, sin considerar la confiabilidad estadística de los resultados, como ya se comentó, dada la autorregresividad de las variables involucradas en la representación y la imposibilidad de asegurar la invariancia temporal de las condiciones que afectan el comportamiento del sistema, la teoría de la probabilidad era considerada como un argumento inapropiado para el análisis de los fenómenos económicos, de modo que, hasta la implementación del proceso de modelización, se utilizaron métodos de estadística matemática "sin probabilidad".

De acuerdo con esta concepción no-probabilística, las primeras interpretaciones atribuyeron la presencia de la componente aleatoria exclusivamente a "errores" en la medición de los factores incluidos en la representación ("*errores en las variables*", según la denominación de Schultz (1925) y Frisch (1934))[26]. Ahora bien, esta interpretación resultaba insuficiente para explicar los errores provenientes de la omisión de variables (en general, de mayor

---

[23]. El cambio de nomenclatura se debe a que, dado el carácter discreto de la información, las representaciones deben ser entendidas como aproximaciones discretas al comportamiento de los fenómenos en un espacio-tiempo continuo-continuo.

[24]. De acuerdo con la tradición el término "modelo" –entendido como estimador de una representación- fue introducido por Frisch en el Congreso de Econometric Society en 1931 y fue utilizado por primera vez por Tinbergen, en 1936 (Magnus; Morgan (1987), Morgan (1989)).

[25]. El desarrollo de métodos de estimación de los coeficientes de representaciones conceptualmente válidas constituyó una cuestión fundamental en la evolución de la econometría.

[26]. La idea de la posibilidad de existencia de errores de medición en las variables fue introducida por Gini (1921).

importancia que los errores en las variables) que requería el agregado de otra componente aleatoria, lo cual implicaba la adopción de métodos estadísticos más avanzados, cuyo desarrollo implicaba una identificación desagregada del origen y las características de los errores. En principio, en el análisis de los ciclos económicos, esta insuficiencia fue resuelta mediante la introducción del concepto de "*errores en las ecuaciones*" y una mayor profundización en el estudio de la naturaleza de la aleatoriedad inherente al comportamiento de los fenómenos.

Originalmente el análisis de los ciclos se debe a Moore (1914)(1923), quien propuso una representación utilizando análisis harmónico y series de Fourier. Alternativamente Persons (1916)(1922-23) desarrolló un método conocido como el "*barómetro de Harvard*", basado en la correlación entre series desfasadas un período (un modelo conocido hoy como "*indicador líder*"). Estos métodos perdieron vigencia ante las críticas de Slutsky (1927) y Yule (1921) (1926)(1927) referidas a la posible falta de significado conceptual de la correlación entre series cronológicas autorregresivas.

La propuesta de Yule se basó en el estudio de series perturbadas. Demostró que un esquema autorregresivo definido por un péndulo sometido a perturbaciones estocásticas genera un proceso "*armónico irregular*" y propuso una forma de representación utilizando cuadrados mínimos en base a una ecuación en-diferencias de la forma $Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2}$[27] más un "shock" aleatorio formado por la combinación de los errores y las perturbaciones que afectaban al comportamiento de $Y_t$. El hecho que los componentes de esta combinación fueran no-separables permite concluir que la solución de Yule no resolvía las cuestiones referidas a la forma en que las perturbaciones eran "absorbidas" por el sistema, ni a su interpretación conceptual.

## 8.- Frisch y la descomposición estructural

Basándose en los trabajos de Wicksell (1907), Hotelling (1927) y Walker (1931), Frisch (1933)(1936) propuso el método de descomposición estructural[28], cuyo objetivo era sistematizar, en ciertas aplicaciones, las investigaciones sobre la estimación de representaciones utilizando series cronológicas mediante métodos de estadística matemática basados en argumentos probabilísticos. Introdujo el concepto de "*shocks erráticos*" para completar la caracterización de las variables $\varepsilon_t$. Las consideró como "*...desvíos del comportamiento estructural (...) como algo nuevo espontáneo agregado a la estructura del proceso*" (p. 408) que dan origen a la modificación aleatoria de las regularidades locales observables en su comportamiento y constituyen "*...la fuente de energía que generan los ciclos económicos*", incorporando de esta forma, parcialmente[29], la teoría de la probabilidad al pensamiento econométrico. Propuso, además, una clasificación de los "shocks" -de acuerdo con la forma en que son absorbidos por el sistema- en "*aberraciones*" y "*estímulos*", definiendo como estímulo a "*... una perturbación cuyo efecto perdura sobre estados sucesivos del sistema*" y como una aberración a "... *una perturbación que influye sobre el proceso solamente en el momento que se produce*" (p. 410), generada por el método de estimación utilizado[30].

---

[27]. Ecuación que puede ser considerada como el origen de las representaciones $AR$ .

[28]. Esta propuesta fue planteada por Frisch como una alternativa crítica al método utilizado habitualmente de la "*descomposición mecánica*" (en las componentes tendencia, estacionalidad, ciclo y variaciones no-sistemáticas).

[29]. Parcialemente en la medida que Frisch no estaba de acuerdo con la aplicación indiscriminada de laoría de la probabilidad en economía.

[30]. En particular, en este caso, como generadas por su método de estimación-identificación, al que denominó "*análisis de confluencia*", cuya finalidad era intentar resolver los problemas generados por la no consideración de relaciones ocultas para el observador, debido a los errores de medición en todas las variables (no solamente en

De acuerdo con la tesis Fechneriana de la "*indeterminación por novedad*"[31] se puede considerar a la postulación de Frisch como el primer intento de interpretación de los estímulos como verdaderas innovaciones[32].

Por otra parte, a partir de los resultados de Slutsky acerca de la capacidad de los procesos aleatorios puros de generar procesos autorregresivos, concluyó "*...que las regularidades observadas en el comportamiento de $Y_t$ podían ser consideradas como derivadas de un caos generado por elementos inconexos, por su propia inconexión*" (Frisch (1936, p. 107)) y clasificó a dichos elementos (a los cuales denominó en forma poco feliz como "*factores caóticamente aleatorios*")[33] en "*coherentes*" e "*incoherentes*" según que fueran o no autorregresivos y demostró -generalizando los trabajos de Yule (1921) sobre "*sumas ponderadas móviles*" de variables aleatorias independientes- que un proceso coherente podía ser aproximado como una combinación lineal de variables incoherentes de la forma:

$$Y_t = \theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \qquad (t = 1,2,\dots)$$

(donde las variables $Y_t$ son centradas y $\{\varepsilon_t\}$ es un ruido blanco.

J. Tinbergen (1935)(1937)(1938)(1939) realizó una notable síntesis de los métodos propuestos por R.A. Fisher y Koopmans y puede ser considerado el continuador inmediato de la obra de Frisch. Si bien coincidió con éste en las propiedades de las representaciones formadas por ecuaciones lineales ponderadas, abandonó su forma estructural combinada, compuesta por ecuaciones diferenciales y en-diferencias, a favor de ecuaciones discretas del tipo de las denominadas "*de rezagos distribuidos*" de la forma[34]:

$$Y_t = a_0 + b_0X_t + b_1X_{t-1} + \varepsilon_t^*$$

(donde los "shocks" $\varepsilon_t^*$ representan los estímulos que preservan la dinámica del sistema).

La generalización del pensamiento probabilístico que generó la propuesta pionera de Frisch -la cual influyó más en los aspectos metodológicos que en los epistemológicos de la teoría de modelos- se vio impulsada por los avances de la teoría de la inferencia, la axiomatización de la teoría de la probabilidad por obra de Kolmogorov (1933) y, posteriormente, por el desarrollo de la teoría de la decisión en condiciones de incertidumbre, la teoría del riesgo, el criterio de optimización mediante la maximización de la utilidad esperada y sus extensiones y por la aparición de la interpretación personalista del concepto de probabilidad por obra de Ramsey y de Finetti (ver Rowley; Hamouda (1987), Hamouda; Rowley (1988))[35]. En este sentido, merecen ser destacados los trabajos de Marschak (1937) -quien propuso la consideración del comportamiento de los operadores económicos como parte de la teoría de los juegos y, en consecuencia, la asimilación de los conceptos de función de preferencias y utilidad.

---

la variable "dependiente", como postulaba el método de regresión utilizado habitualmente) y por la multicolinealidad (ver Hendry; Morgan (1989)).

[31]. Fechner (1866)(1871)(1906). Esta tesis se basa en el principio que en la evolución de los fenómenos se van generando nuevas condiciones iniciales las cuales, de acuerdo con una ley general, conducen a efectos no ocurridos previamente y, por lo tanto, no permiten la recurrencia de comportamientos idénticos. Pero que, no obstante, cada conjunto de dichas condiciones conserva alguna similitud con los anteriores por lo que, si bien todo fenómeno posee cierta dosis de aleatoriedad objetiva, su comportamiento no es completamente libre. Ver Landro (2010).

[32]. Frisch (1938): "*...los estímulos pueden ser considerados como condicionantes de la evolución posterior del fenómeno, es decir como causantes de una suerte de cambio permanente de sus condiciones iniciales*" (p. 408).

[33]. La calificación de "poco feliz" se basa en la aparente asimilación de los conceptos de caótico y aleatorio, de complejidad y aleatoriedad.

[34]. El concepto de "*rezagos distribuidos*" fue introducido por I. Fisher en 1920, en un trabajo sobre oferta monetaria (ver Alt (1942)).

[35]. Ver Landro; González (2013)(2014).

## 9.- Haavelmo y la revolución probabilística

La culminación de la aproximación probabilística a la econometría se produjo con la llamada "*revolución Haavelmiana*" (Morgan (1987)(1989)). A partir de la propuesta esencialmente estocástica de Slutsky, de la postulación de Moore (1914) -reconsiderada por Schultz (1930)(1939)- sobre la utilización del análisis harmónico y de la aplicación rigurosa de la teoría de la probabilidad por Koopmans (1937), Haavelmo (1938)(1939)(1940a)(1940b) (1943)(1944) postuló que las ecuaciones estructurales debían ser interpretadas como leyes de comportamiento sólo en un sentido estadístico y que los modelos econométricos debían asumir la forma de las "*ecuaciones estructurales estocásticas*" de Frisch, incluyendo "*coeficientes estructurales estocásticos*". Esto lo condujo a la conclusión que la formalización de un procedimiento de modelización estocástica consistente y generalizado (basado en observaciones no-repetibles) requería inevitablemente de los métodos de máxima verosimilitud de Fisher y de la teoría de los tests de hipótesis de Neyman-Pearson, que permitieran decidir acerca de la correspondencia entre la interdependencia de las variables económicas y su tratamiento estadístico[36]. Haavelmo recurrió así al concepto de distribución conjunta de todas las variables observables como un argumento para justificar la forma de dicha interdependencia estadística y concluyó que para que esta solución fuera posible, era necesario atribuirle a los términos representativos de los errores ciertas distribuciones de probabilidades.

Esta forma de especificación mediante la definición de la distribución de probabilidades conjunta es considerada, en general, como el fundamento de la revolución probabilística Haavelmiana y de los modelos de ecuaciones simultáneas y determinó la diferencia "*...entre la construcción de modelos econométricos y la formulación matemática de la teoría económica*" (Qin (1993, p. 58)). Es decir determinó la diferencia entre los modelos econométricos y los modelos económicos.

Inmediatamente Mann; Wald (1943) aplicaron la especificación mediante la definición de la distribución conjunta y demostraron la consistencia y la Normalidad asintótica de los estimadores máximo-verosímiles en las ecuaciones lineales de la forma:

$$a_0 Y_t + a_1 Y_{t-1} + \cdots + a_j Y_{t-j} = \varepsilon_t$$

(donde $\{Y_t\}: I(0)$ y $\varepsilon_t$ denota un vector de perturbaciones no-autocorrelacionadas, idénticamente distribuidas y con momentos de orden superior finitos).

Como corolario de esta aplicación surgió el problema práctico de cómo asegurar la completitud del sistema de ecuaciones desde un punto de vista estadístico. Con este fin Koopmans (1950) redefinió el concepto de completitud de Tinbergen y, a partir de la clasificación de las variables en exógenas y endógenas, postuló que un modelo se puede considerar completo cuando el número de ecuaciones coincide con el número de variables endógenas del sistema. Definió como variable exógena aquella que no está afectada por las variables endógenas ni por las perturbaciones e introdujo el concepto de variable predeterminada que incluye a las variables desfasadas y a las variables exógenas correspondientes al período $t$. A partir de estas definiciones propuso una transformación del método de identificación basado en la distribución conjunta de las perturbaciones, en otro basado en el producto de la distribución de las perturbaciones condicionada por las

[36]. En general la literatura considera que el punto de partida de la revolución Haavelmiana se encuentra en la respuesta de Haavelmo (1943) (fundamentalmente en Schumpeter (1939) y en el debate Keynes-Tinbergen) a los prejuicios planteados con respecto a la naturaleza no-experimental de las observaciones económicas y, en consecuencia, a la capacidad de la inferencia estadística basada en la probabilidad para verificar las supuestas verdaderas relaciones propuestas por la teoría económica.

variables exógenas multiplicada por la distribución marginal de dichas variables exógenas.

Como una excepción a esa corriente de pensamiento que consideraba a la teoría de la probabilidad exclusivamente como el fundamento de los métodos estadísticos a aplicar en la construcción de los modelos estructurales, entendidos estos como representaciones (incompletas) de las verdaderas trayectorias de los fenómenos económicos más una componente representativa de los errores aleatorios, cabe mencionar los trabajos de Wold y Tintner.

**10.- Wold y la descomposición predictiva**

El teorema de la descomposición predictiva de Wold (1938) -su tesis doctoral- resultó el hito fundamental en el análisis del tratamiento causal moderno de los fenómenos dinámicos. Consideró a las series cronológicas como la porción observable de un proceso estocástico, es decir, como "*una realización de una distribución de infinitas dimensiones*" (p. 4) y completó las propuestas de Yule y Slutsky demostrando que todo proceso $\{Y_t\}: I(0)$ puede ser desagregado en una componente estructural aproximable por un $AR(p)$ $(p \gg 0)$ y una combinación lineal de infinitos "*errores*" incorrelacionados, asimilables a variables (latentes) no-observables:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

(donde las variables $Y_t$ son centradas).

Cabe destacar que más allá que la condición de aleatoriedad pura implicaba suponer que las variables $\varepsilon_t$ se distribuían de acuerdo con una función de densidad simétrica con valor esperado nulo y la condición de estacionariedad implicaba suponer que su varianza era finita e, ignorando la aproximación asintótica a la Normalidad que proporcionaba la versión de Laplace del teorema central del límite, Wold consideró "*insoportablemente arbitrario*" (Epstein (1987), p. 161) asignar una distribución de probabilidades particular a las variables $\varepsilon_t$.

Como corolario de este trabajo liminar sobre la teoría pura de los procesos de parámetro discreto y del análisis de regresión de H. Cramér y en oposición a los modelos de ecuaciones simultáneas, Wold desarrolló una aproximación recursiva de la forma[37]:

$$Y_t^{(i)} = F\left(Y_{t-1}^{(i)}, \dots, Y_{t-1}^{(i-1)}, \dots, Y_{t-2}^{(i)}, \dots, Y_{t-1}^{(n)}, Y_{t-2}^{(n)}, \dots\right) + \varepsilon_t^{(i)}$$

(Bentzel; Wold (1946)) y demostró (Wold (1949)(1954)) que este modelo proporcionaba una interpretación causal y que, en general, todo conjunto de series cronológicas puede ser representado formalmente como un sistema de encadenamiento causal ((1948)(1951))[38].

La importancia fundamental de estos resultados está evidenciada en el hecho que todo el desarrollo posterior de la teoría econométrica de los fenómenos dinámicos se basó en corolarios y generalizaciones del teorema de Wold.

**11.- Tintner y el método de la diferencia variable**

Contrariamente a la posición de Wold -quien justificó la utilización de la teoría pura de la

[37]. Con respecto a esta ecuación, debe tenenerse en cuenta que, en el planteo dinámico de Wold las relaciones simultáneas no son justificables en la medida que no permiten establecer el sentido de la causalidad.
[38]. Wold (como un producto de la posición adoptada por la escuela sueca de economía) rechazó el paradigma Mashalliano-Walrasiano del equilibrio general como un estado persistente de la dinámica económica.

probabilidad en las representaciones de fenómenos económicos basándose en elementos abstractos inherentes a la definición clásica, Tintner (1938a)(1938b) planteó la posibilidad de utilizar a la probabilidad, de acuerdo con una interpretación más amplia y más flexible, como un argumento para vincular "*...la teoría económica de las expectativas y la teoría estadística de los errores*" (1938b, p. 154)[39]. Consideró a los errores aleatorios como generados por la imposibilidad de obtener "*...las representaciones óptimas de los factores controlables*" y de predecir "*...los factores no-controlables*", clasificándolos como errores en las variables y errores debidos a la influencia de fluctuaciones aleatorias en la descomposición de los sistemas de series cronológicas y postuló que los primeros debían ser tratados por el método de la diferencia variable y los segundos por los métodos postulados por Wold. En particular, su propuesta se basó en la interpretación Carnapiana, la cual admite dos conceptos posibles de probabilidad: $P_{r_1}$ relacionado con el "*grado de confirmación*" y $P_{r_2}$, relacionado con la "*frecuencia empírica*" y concluyó que la primera definición podía constituir el fundamento de una teoría de los tests de hipótesis apropiada para la econometría[40].

## 12.- Marschak, Muth y la autocorrelación entre las perturbaciones

En desacuerdo con la propuesta de Wold, según la cual $\{\varepsilon_t\}: WN - débil$, Marschak (1953) justificó la posible presencia de autocorrelaciones entre las perturbaciones y postuló que"*...deben ser consideraradas como parte del comportamiento estructural, en la medida que no hay ninguna razón económica para descartar la posibilidad de que formen un proceso estocástico en el que cada shock depende de uno o más de sus predecesores*" (p. 21). Esta propuesta hizo que la investigación econométrica se dirigiera (fundamentalmente por obra de Cochran; Orcutt (1949), Durbin (1960), Sargan (1959)(1961)) al estudio de las implicaciones producidas por las autocorrelaciones entre las perturbaciones sobre las propiedades de los estimadores de los coeficientes estructurales.

Muth, en su trabajo fundamental de 1961, propuso un retorno a la interpretación de Frisch (1933) de los estímulos en el comportamiento estructural de un sistema como una caracterización de los efectos dinámicos de los "shocks" aleatorios autocorrelacionados (pero económicamente no especificados), débilmente estacionarios de segundo orden, asimilables a variables exógenas y representables como una combinación lineal de variables $\varepsilon_t$ no-autocorrelacionadas con valor esperado nulo y varianza constante ($Y_t = f[\Omega^*(Y_t)] + \varepsilon_t^*$ donde $\{\varepsilon_t\}: WN$)[41].

## 13.- Theil y el transcurso hacia el aleatorismo

En este punto la hipótesis de la ortodoxia econométrica original, que sostenía que siempre era posible especificar un modelo teórico que explicara la verdadera ley que rige el comportamiento de un fenómeno[42], comenzó a debilitarse y surgió la necesidad de

---

[39]. Debe tenerse en cuenta que, dado que Wold no consideró la necesidad de testear las representaciones propuestas respecto de su adecuacidad a las correspondientes posiciones teóricas, no reparó en las condiciónes de aplicabilidad del concepto de probabilidad. Su propuesta postulaba que los sistemas recursivos no estaban afectados por los problemas de identificación.
[40]. Ver Landro (2010).
[41].A diferencia de Frisch (1937) -quien utilizó una representación $MA$ para descomponer una variable económica- Muth utilizó una representación $MA$ para descomponer un "shock" aleatorio.
[42]. Entendida la especificación como "*...la selección de la forma matemática de la población*" (Koopmans (1937, p. 3)).

modificaciones "ad hoc" en las especificaciones de los modelos estructurales para representar las eventuales regularidades locales, de interpretar a los modelos como hipótesis generales sobre las principales relaciones causales sugeridas por la teoría económica, testeables a partir de las observaciones. En este contexto de un naciente paradigma aleatorista, Theil (1957)(1958), a partir del supuesto que "*...los modelistas en general no conocen la verdadera especificación de una representación*" (Theil (1958, p. 215)) (cabría agregar, suponiendo que ésta exista), propuso una definición alternativa de "*especificación*" y, por lo tanto, de capacidad de predicción, basada en la minimización de la varianza residual como criterio de optimización para la selección de un modelo[43], pero sin asignarle ninguna interpretación conceptual a los residuos.

En general en el paradigma estructuralista los principales esfuerzos estuvieron dirigidos a la identificación y estimación de los coeficientes estructurales a partir de las formas reducidas, dedicándole muy poca atención a la verificación del cumplimiento de las condiciones impuestas a los "shocks" exógenos (Granger; Newbold (1977): "*Los residuos son tratados habitualmente por los econometristas como meras molestias de poca importancia. Muchos textos de econometría introducen a los modelos como un conjunto de relaciones determinísticas entre variables y adicionan descuidadamente términos representativos del 'error' a las ecuaciones para explicar cosas como errores en la especificación del modelo y en la medición de las variables. La utilización de denominaciones como 'residuos' y 'errores' implica juicios de valor sobre la importancia de dichos términos. Si bien 'error' es una denominación adecuada en el contexto de la predicción, una denominación más apropiada podría ser 'innovaciones'*'" (p. 8)).

Un breve retorno al determinismo en este transcurso hacia el aleatorismo lo constituye la obra de Leamer (1978), quien reinterpretó la propuesta de Theil sobre el problema de la especificación desde una perspectiva Bayesiana[44]. Consideró a las leyes económicas como el fundamento teórico de las representaciones estructurales y propuso considerar a los errores como independientes de esa "verdadera" formulación teórica generados por la agregación de elementos "*no-observables*" pero "*manejables*", debidos exclusivamente a deficiencias en la especificación de la representación[45] y concluyó, en consecuencia, que *una representación "completa" implicaría la eliminación de las perturbaciones*[46].

### 14.- Sargent, Sims y los modelos de series cronológicas

Dadas las debilidades evidenciadas por el estructuralismo en la formulación dinámica de las representaciones y a partir de extensiones del teorema de Wold, Sargent (1976)(1977), Sargent; Sims (1977) y Sims (1980) impulsaron una aproximación a la explicación del comportamiento de los fenómenos económicos mediante modelos puros de series cronológicas. En particular, propusieron la utilización de representaciones de la forma $\Phi(B)Y_t = \varepsilon_t$, en las cuales el orden $p$ del operador $\Phi(B)$ fuera tal que $\{\varepsilon_t\}$ pudiera ser considerado un proceso de innovaciones no-autocorrelacionadas, de modo que esta

---

[43]. Es necesario tener en cuenta que "*...dado que este criterio no siempre es concluyente ni factible, deben considerarse dos criterios adicionales de naturaleza más subjetiva: plausibilidad y simplicidad*" (Theil (1958, p. 208)).

[44]. La diferencia fundamental entre Theil y los econometristas clásicos y Leamer y los econometristas Bayesianos radica en que estos estaban más interesados en el análisis de las propiedades de los estimadores de los coeficientes que en la minimización de la varianza residual de los clásicos.

[45]. Ver Qin; Gilbert (2001).

[46]. Una interpretación que puede considerarse resumida en la expresión atribuida a Tukkey, según la cual "*El hombre construye* $f\left[\Omega^*(Y_t)\right]$ *y Dios nos proporciona* $\varepsilon_t$ ".

condición de $AR(0)$ del $\{\varepsilon_t\}$ proceso se verificara por construcción y no por una hipótesis "a priori" y la representación $\Phi(B)Y_t = \varepsilon_t$ constituyera una caracterización económicamente válida del comportamiento de $Y_t$.

La representación inversa, $Y_t = [\Phi(\mathrm{B})]^{-1}\varepsilon_t$ retorna, en una interpretación aleatorista, a la propuesta de Slutsky-Wold de representación lineal de un proceso "*coherente*" ($\{Y_t\}$) en términos de un proceso "*incoherente*" ($\{\varepsilon_t\}$). En la que el operador $[\Phi(\mathrm{B})]^{-1}$, como función que describe la medida del impacto de los "shocks" aleatorios, constituye el argumento más importante para considerar a las perturbaciones como innovaciones y comprender cómo estas innovaciones son asimiladas por el proceso $\{Y_t\}$.

## 15.- Conclusiones

La primera conclusión a la que permiter arribar esta escueta reseña sobre el desarrollo de la teoría de la probabilidad y su influencia en la explicación del comportamiento de los fenómenos económicos, es que el origen de la teoría de los modelos estocásticos dinámicos se remonta a una época muy anterior a la que propone, en general, la literatura econométrica. Que su génesis puede considerarse asociada: **i)** a los postulados del teorema de la inversión de la probabilidad de J. Bernoulli (1713), que planteó la necesidad de hallar el nexo entre las probabilidades "a priori" (definidas a partir de un razonamiento que va de las causas a los efectos) y las probabilidades "a posteriori" (definidas a partir de un razonamiento que va de los efectos a las causas) y, en particular, **ii)** a la novedad conceptual de la solución propuesta por Simpson, relacionada con la definición de una distribución de probabilidades de los errores de observación aleatorios y la obtención, por añadidura, de la inversión de la probabilidad.

La solución de Simpson condujo al criterio de los cuadrados mínimos de Legendre el cual, provisto de una interpretación probabilística tuvo su culminación en la síntesis de Gauss-Laplace.

La adopción de este criterio de optimización basado en la minimización de la suma de los cuadrados de los errores aleatorios y en la Normalidad de su distribución, y el surgimiento de un nuevo método de inferencia estructurado a partir de conjuntos de observaciones no homogéneas, dio origen a la aplicación de la teoría de los modelos estocásticos a la econometría.

A partir de los postulados de la versión de Laplace del teorema central del límite y de la asimilación de fenómenos dinámicos a procesos estocásticos la econometría evolucionó adoptando, alternativamente, interpretaciones determinísticas y aleatoristas de las perturbaciones.

El análisis de la interpretación econométrica del concepto de error aleatorio comenzó con Yule, quien propuso una forma de representación a partir de un esquema autorregresivo e introdujo el concepto de "shock" aleatorio formado por la combinación de errores y perturbaciones que afectaban el comportamiento del sistema. Pero, que los componentes de esta combinación fueran no-separables, permitió arribar a una segunda conclusión: que esta solución no logró resolver el problema de la interpretación conceptual de las perturbaciones ni a la forma en que eran "absorbidas" por el sistema

Fue Frisch quien propuso una primera solución a esta cuestión considerando a los "shocks" como "*algo nuevo y espontáneo agregado a la estructura del proceso*", que da origen a la modificación aleatoria de las regularidades locales observables y constituye "*...la fuente de* *energía que generan los ciclos económicos*", incorporando de esta forma, parcialmente, la teoría de la probabilidad al pensamiento econométrico. Esta interpretación condujo a una tercera conclusión: que, de acuerdo con la tesis de la "*indeterminación por novedad*" se puede considerar a la postulación de Frisch como el primer intento de interpretación de los "estímulos" como verdaderas innovaciones Fechnerianas.

En oposición a la propuesta de Haavelmo referida a la construcción de modelos de ecuaciones simultáneas, cabe mencionar los trabajos de Wold y Tintner, basados en un sistema de encadenamiento causal. En particular, el teorema de la descomposición predictiva de Wold permitió arribar a una cuarta conclusión: que esta propuesta constituye la piedra angular sobre la que se construyó todo el análisis del tratamiento moderno de los fenómenos dinámicos, cuyas generalizaciones generaron todo el desarrollo posterior de la teoría econométrica.

La quinta y última ("*last but not least*") conclusión a la que se arribó es: **i)** que, diez años después de la aparición de la propuesta de Wold, el progresivo debilitamiento de la hipótesis determinista de la ortodoxia econométrica generó la necesidad de introducir modificaciones "ad hoc" en las especificaciones de los modelos estructurales para representar las eventuales regularidades locales, y de interpretar a los modelos como hipótesis generales sobre las principales relaciones causales sugeridas por la teoría económica, testeables a partir de las observaciones y **ii)** que este cambio en la teoría de modelos hacia un paradigma aleatorista se fundamenta en la obra de H. Theil.

**Bibliografía**


Adrain, R. (1809): "Research concerning the probabilities of the errors wich happen in making observations". The analyst, vol. 1 (93-109). Reeditado en Stigler, S., 1980.

Alt, F.L. (1942): "Distributed lags". Econometrica, vol. 10 (113-128).

Arnauld, A.; Nicole, P. (1662): "*La logique ou l'art de penser*". Reeditado por Flammarion, 1970.

Baxter, D. (1870): "National income and taxation of the United Kingdom". Congrés International de Statistique à la Haye (137-144).

Bentzel, R.; Wold, H. (1946): "On statistical demand analysis from the viewpoint of simultaneous equations". Skandinavisk Aktuarietidskrift, vol. 29 (95-114).

Bernoulli, D. (1778): "Diiudicatio maxime probabilis plurium observationem discrepatium atque verosimillima induction inde formanda". Acta Academiæ Scientiarum Imperialis Petropolianæ. Reeditado en "*Die Werke von Daniel Bernoulli*", Birkhäuser.

Bernoulli, J. (1975): "*Meditationes*" (1684-1690). En "*Die Werke von Jakob Bernoulli*". Birkhäuser, 1975.

Bernoulli, J. (1713): "*Ars conjectandi*". Thurnisiorum.

Bernoulli, N. (1709): "*De usu artis conjectandi in jure*". Basilea. Reeditado en "*Die Werke von Jakob Bernoulli*". Birkhäuser, 1975.

Boscovich, R.J.; Maire, C. (1755): "*De Litteraria Expeditione per Pontificiam ditionem ad dimetiendas duas Meridiani gradus*", Palladis.

Bowley, A.L. (1921): “Elements of statistics”. P.S. King.

Cochran, D.; Orcutt, G. (1949): “Application of least squares regression to relationships containing autocorrelated error terms”. JASA, vol. 44 (32-61).

Côtes, R. (1722): “Opera miscellanea sive æstimatio errorum in mixta mathesi, per variationes partium trianguli plani et sphærici”. En Smith, R. (ed.).

Cournot, A.A. (1838): “*Recherches sur les principles mathématiques de la théorie des richesses*”. París.

Cournot, A.A. (1843): “*Exposition de la théorie des chances et des probabilités*” Hachette. Reeditado en “*Cournot, A.A. OEvres completes*”. Vrin, 1973-1984.

Dale, A.I. (1991): “*A history of inverse probability: From Thomas Bayes to Karl Pearson*”. Springer.

Daston, L. (1988): “*Classical probability in the enlightenment*”. Princeton University Press.

de Moivre, S.D. (1718): “*The doctrine of chances: or a method of calculating the probability of events in play*”. W.- Pearson.

de Moivre, S.D. (1730): “*Miscellanea analítica de seriebus et quadraturis*”. Tonson & Watts.

de Moivre, S.D. (1733): “Approximatio ad summam terminorum binomii $(a+b)^n$ in seriem expansi”. Reproducción en Archibald, R.C. (1926).

Drobisch, M.W. (1871): “Über Mittelgrössen und die Anwendberkeit derselben auf die berechnung des Steigens und Sinkens des geldwertes”. Jarbücher für Nationalökonomie uns Statistik (1871).

Durbin, J. (1960a): “The fitting of time-series models”. Review of the International Statistical Institute, vol. 28 (233-243).

Durbin, J. (1960b): “Estimation of parameters in time-series regression models”. JRSS, Serie B, vol. 22 (139-153).

Edgeworth, F.Y. (1881): “*Mathematical psychics: An essay on the application of mathematics to the moral sciences*”. C. Keegan. Reeditado por A.M. Kelley, 1967.

Edgeworth, F.Y. (1896): “Eine Erwiderung”. Jahrbücher für Nationalökonomie und Statistik, vol. 12 (838-845).

Edgeworth, F.Y. (1910): “On the applicacion of the calculus of probabilities to statiscs”. Bulletin de l’Innstitut Internationale de Statistique, vol. 18 (505-537).

Epstein, R.J. (1987): “*A history of econometrics*”. North-Holland.

Euler, L. (1749): “*Recherches sur la question des inegalités du mouvement de Saturneet et Jupiter, sujet propose por le prix de l’aneé, par l’Academie Royale des Sciences de Paris*”. Turici.

Fechner, G.T. (1866): “*Elemente der Psycophysics*”. Versión en inglés, Holt-Rinehart-Winston (1966).

Fechner, G.T. (1871): “Zúr experimentalen Ästhetik”. Abhandlungen der Könglich säschischen gesellchaft der Wissenschaften, vol. 9.

Fechner, G.T. (1906): "*Send-Avesta, oder über die Dinge des Himmels und des Jenseits*". 3ra. edición, Leopol Voss.

Fellner, F. (1901): "L'evaluation de la richesse nationale". Bulletin de l'Institute International de Statistique, vol. 13, part 2 (96-136).

Fellner, F. (1905): "Die Scätzung der Volkseinkomments". Bulletin de l'Institute International de Statistique, vol. 14, part 3 (109-120).

Fisher, I. (1892): "Mathematical investigations in the theory of value and prices". Kelley.

Fisher, I. (1911): "The purchasing power of money". Kelley.

Fisher, I. (1922): "The making of index numbers: A study of the varieties, tests and reliability". Kelley.

Frisch, R. (1933): "Propagation problems and impulse problems in dynamic economics". En "*Economic essays in honour of Gustav Cassel*". Allen-Unwin.

Frisch, R. (1934): "*Statistical confluence analysis by means of complete regression systems*".Universitets Økonomiske Institutt.

Frisch, R. (1936): "Time series and business cycle analysis: Economic macro dynamics". En "Report of the work done under the direction of Professor I. Wedervang, at the U. Institute of Economics, Oslo". Enero 1932-Junio 1936, Rockefeller Archive Centre.

Frisch, R. (1938): "Statistical versus theoretical relations in economic macrodynamics". Memorandum, Oslo. Reproducido en Hendry; Morgan (1995).

Galilei, G. (1632): "*Dialogo sui massimi sistema del mondo*". En "*Opere*", editado por A. Favaro, 1968.

Gauss, C.F. (1809): "*Theoria motus corporum coelestiorum*" Partes et Besser. Reimpreso por Dover, 1963.

Gauss, C.F. (1816): "Bestimmung dert Genauigkeit der Beobachtungen". Reeditado en "*Carl Friederich Gauss*". Königlichen Gesellschaft der Wissenschaften, 1880.

Gauss, C.F. (1823): "*Theoria combinationis observationum erroribus minimis obnoxiæ*". Dietrich.

Gini, C. (1921): "Sull' interpolazione di una retta quando i valori della variable indipendente sono affetti da errori accidentali". Metron, vol. 1 (63-82).

González, M.L.; Landro, A.H. (2013): "Acerca de la interpretación del concepto de perturbación en los procesos discretos de parámetro finito". IV Seminario de Docencia, Investigación y Transferencia en las Cátedras de Martemáticas para Economistas.

Gowing, R. (1983): "*Roger Cotes-Natural philosopher*". Canbridge University Press.

Granger, C.W.J.; Newbold, P. (1977): "*Forecasting economic time series*". Academic Press.

Hagstroem, K, -G. (1938): "Pure economics as a stochatical theory". Econometrica, vol. 6 (40-47).

Haavelmo, T. (1938): "The method of supplementary confluent relations. Illustrated by a study of stock prices". Econometrica, vol. 6 (203-218).

Haavelmo, T. (1939): “Statistical testing of dynamic systems if the series observed are shock cumulants”. Report of the 5th Annual Reseaerch Conference on Economics and Statistics.

Haavelmo, T. (1940a): “The inadequacy of testing dynanic theory by comparing the theoretical solutions and observed cycles”. Econometrica, vol. 8 (312-321).

Haavelmo, T. (1940b): “The problem of testing economic theories by means of passive observations”. Report of the 6th Annual Reseaerch Conference on Economics and Statistics.

Haavelmo, T. (1943): “The statistical implications of a system of simultaneous equations”. Econometrica, vol. 11 (1-12).

Haavelmo, T. (1944): “*The probability approach in econometrics*”. Suplemento a Econometrica, vol. 12.

Halley, E. (1693): “An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslau wit an attempt to escertyain the price of annuities upon lives”. Philosophical Trensactions of the Royal Society, vol. 17 (596-610).

Hamouda, O.; Rowley, J.C.R. (1988): “*Expectation, equilibrium and dynamics*”. Hemel Hempstead.

Hendry, D.F.; Morgan, M.S. (1989): “A re-analysis of confluence analysis”. Oxford Economic Journal, vol. 41 (35-52).

Hotelling, H. (1927): “Differential equations subjet to error”. JASA, vol. 22 (283-314).

Hume, D. (1718): “*An inquiry concerninghuman understanding*”. Reeditado por Handel (1955).

Jevons, W.S. (1871): “*The theory of practical economy*”. MacMillan.

Kendall, M.G.; Plackett, R.L. (eds.)(1977): “*Studies in the history of statistics and probability*”. Griffin.

Keynes, J.M. (1921): “*A treatise of probability*”. MacMillan, 1963.

Keynes, J.M. (1993): “*The general theory of employment, interest and money*”. En “*Collected writings of John Maynard Keynes*”. MacMillan, Cambridge University Press for the Royal Economic Society.

Kolmogorov, A.N. (1933): “*Grundbegriffe der Wahrcheinlichkeitsrechnung*”. Springer. Traducción al inglés: “*Foundations of the theory of probability*”, Chelsea, 1950.

Koopmans, T.C. (1937): “*Linear regression analysis of economic time series*”. Netherlands Economic Institute.

Koopmans, T.C. (1950): “*Statistical inference in dynamic economic models*”. Cowles Commission, Monografía nº 10.

Kops, (1869): “Revenu annuel de la nation”

Lambert, J.H. (1760): “Photometria, sice de mensura et gradibus luminis, colorum et umbræ”. Detleffsen.

Lambert, J.H. (1765-1772): “Beyträge zum Gebrauche der Mathematik und deren Anwendung”. Verlage des Buchladens des Realschule. Vols. 1-3.

Landro, A.H. (2010): "*Acerca de la probabilidad*". Ediciones Cooperativas.

Landro, A.H.; González, M.L. (2013): "Acerca de la interpretación económica de los conceptos de esperanza matemática y esperanza moral y sus contraejemplos. Parte I: Las soluciones clásicas". XIII Jornadas de Tecnología Aplicada a la Educación Matemática Universitaria.

Landro, A.H.; González, M.L. (2014): "Acerca de la interpretación económica de los conceptos de esperanza matemática y esperanza moral y sus contraejemplos. Parte II: Las soluciones frecuencista, logicista y subjetivista". IV Seminario de Docencia, Investigación y Transferencia en las Cátedras de Martemática para Economistas.

Laplace, P.S. (1774): "Mémoire sur la probabilité des causes par les événements". Mémoires de l'Académie Royal des Sciences Presentés. Savants ètrangers., vol. 6 (621-656).

Laplace, P.S. (1781): "Mémoire sur les probabilités". Mémoires de l'Académie Royal des Sciences de Paris (227-332). Acta academica Scientiarum Imperialis Petropolitanæ (3-23). Traducción al inglés en Kendall (1961)(3-13).

Laplace, P.S. (1786): "Mémoire sur les approximations des formules qui sont fonctions de trés grans nombres". Mémoires de l'Académie Royal des Sciences de Paris (1-88).

Laplace, P.S. (1811): "Mémoire sur les intégrales définies et leur application aux probabilités, et spécialment á larecherche du millieu qu'il faut choisir entre les résultats des observations". Mémoires de l'Académie Royal des Sciences de Paris, Primera Serie, vol. 11 (279-437).

Laspeyres, A.Th. (1864a): "Hamburger Warenpreise und die californisch-australischen Goldentdeckungen seit 1848". Jahrbücher für Nationalökonomie und Statistik (81-220).

Leamer, E.E. (1978): "*Specification searches: Ad hoc inference with non-experimental data*". Wiley.

Legendre, A.M. (1805): "Nouvelles methods pour la determination des orbites des comètes". Firmin Didot.

Leibniz, G.W. (1678): "*De incerti æstimatiome*". Berlin.

Magnus, J.R-; Morgan, M.S. (1987): "The ET interview: Professor J.Tinbergen". Econometric Theory, vol. 4 (187-209).

Mann, H.B.; Wald, A. (1943): "On the statistical treatment of linear stochastic difference equartions". Econometrica, vol. 11 (173-220).

Marschak, J. (1937): "Utility and probability in human choice". En Report of 3rd. Annual Research Conference on Economics and Statistics. Cowles Commission.

Marschak, J. (1953): "Economic measurements for policy and prediction". En Hood; Koopmans (eds.).

Mayer, T. (1750): "Abhandlung über die Umwalzung des Monas um seine Axe und die Scheinbare Bewegung der Mondsflecten"

Moore, H.L. (1914): "*Economic cycles: Their law and cause*". Nueva York.

Moore, H.L. (1923): "*Generating economic cycles*". Nueva York.

Morgan, M.S. (1987): "Statistics without probability and Haavelmo´s revolution in econometrics". En Krüger; Gingerezer; Morgan (eds.).

Morgan, M.S. (1989): "*The history of econometric ideas*". Cambridge University Press.

Paasche, H. (1871): "Über die Preisentwicklung der letzen Jahre nach den Hamburger Börsenentwicklungen". Jahrbücher für Nationalökonomie und Statistik.

Pearson, K. (1893a): "Asymetrical frequency curves". Nature, vol. 48 (20-35).

Pearson, K. (1893b): "Contributions to the mathematical theory of evolution, I". Proceedings of the Royal Society, vol. 54 (211-242).

Pearson, K. (1895): "Contributions to the mathematical theory of evolution, II". Philosophical Transactions of the Royal Society, Sereie A, vol. 186 (343-414).

Pearson, K. (1896): "Contributions to the mathematical theory of evolution, III". Philosophical Transactions of the Royal Society, Sereie A, vol. 187 (174-227).

Pearson, K. (1896): "*The history of statistics in the 17th and 18th centuries, against the changing background of intellectual scientific and religious thought. Lectures given at the University College of London, during the academis sessions 1921-1923*". Griffin.

Persons, W.M. (1916): "Construction of a business barometer". American Economic Review, vol. 6 (739-769).

Persons, W.M. (1922-23): "Correlation of time series". JASA, vol. 18 (713-726).

Poisson, S.D. (1837): "Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédés des règles générales du calcul des probabilités. Bachelier.

Prigogine, I.; Nicolis, G. (1977): "*Self-organizatgion in non-equilibrium system, from dissipative structures to order to fluctuztions*". Wiley.

Qin, D. (1993): "*The formation of econometrics. A historical perspective*". Clarendon.

Qin, D.; Gilbert, C.L. (2001): "The error term in the history of time series econometric". Econometric Theory, vol. 17 (424-450).

Quetelet, A. (1848): "Sur la statistique morale et les príncipes qui donent en former la base". Nouveaux Mémoires de l'Académie Royal des Sciences et Belles Lettres de Belgique, vol. 21 (67-114).

Rowley, J.C.R.; Hamouda, O. (1987): "Troublesome probability and economics". Journal of Post Keynesian Economics, vol. 10 (44-64).

Sargan, J.D. (1959): "The estimation of relationships with autocorrelated residuals by the use of instrumental variables". JRSS, Serie B, vol. 21 (91-105).

Sargan, J.D. (1961): "The maximum likelihood estimation of economic relationships with autorregressive residuals". Econometrica, vol. 29 (414-426).

Sargent, T.J. (1976): "A classicalmacroeconomic model for the United States". Journal of Political Economy, vol. 84 (207-237).

Sargent, T.J. (1977): "The demand for money during hyperinflations under rational expectations". International Economic Review, vol. 18 (59-82).

Sargent, T.J. ; Sims, C.A. (1977): "Business cycle modelling without pretending to have too much 'a priori' economic theory". En Sims, C.A. (ed.).

Schultz, H. (1925): “The statistical law of demand”. Journal of Political Economy, v ol. 33 (481-504 y 577-637).

Schultz, H. (1930): “The meaning of statistical demand curves”. Veröffentlichungen der Frankfurter Gessellschaft für Konjunkturforaschung, Leipzig.

Schultz, H. (1939): “*The theory and measurement of demand*”. Chicago.

Schumpeter, J. (1939): “*Business cycles*”. Nueva York.

Sent, E,-M. (1998): “*The evolving rationality of rational expectations: An assessment of Thomas Sargent’s achievements*”. Cambridge University Press.

Simpson, T. (1755): “On the advantage of taking the mean of a number of observations in practical astronomy. A letter to the Right Honorable George Earl of Macclesfield. President of the Royal Society”. Philosophical Transactions of the Royal Society, vol. 49 (82-93).

Simpson, T. (1757a): “*Miscellaneous tracts on some curious, and very interesting subjects in maechanics, physical-astronomy and speculative mathematics*”. J. Nourse.

Simpson, T. (1757b): “An attempt to show the advantage arising by taking ther mean of number of observations in practical astronomy”. En Simpson, T. (1757).

Sims, C.A. (ed.)(1977): “*New methods in business cycle research*”. Federal Reseerve Bank of Minneapolis.

Sims, C.A. (1980): “Macroeconomic and reality”. Econometrica, vol. 48 (1-48).

Slutsky, E. (1937): “The summation of random causes as the source of cyclic processes”. Econometrica, vol. 5 (105-146). Original en ruso, 1927.

Smith, R. (ed.) : “*Opera miscellanea*”. Cambridge, 1722.

Staehle, H. (1933): “Henry L. Moore and statistical economics”. Econometrica, vol. 1 (73-86).

Stigler, S.M. (1975): “Napoleonic statistics: The work of Laplace”. Biometrika, vol. 62

Stigler, S.M. (1986a): “Laplace 1774 memoir on inverse probability”. Statistical Science, vo. 1 (359-378).

Stigler, S.M. (1986b): “*The history of statistics: The measurement of uncertainty before 1900*”. Harvard University Press.

Theil, H. (1957): “Specification errors and the estimation of economic relationships”. Review of International Statistical Institute, vol. 25 (41-51).

Theil, H. (1958): “*Economic forecast and policy*”. North-Holland.

Tinbergen, J. (1935): “Annual survey: Suggestion on quantitatyive business cycle theory”. Econometrica, vol. 3 (241-308).

Tinbergen, J. (1936): “*Grondproblem der Theoretische Statistiek*”. Bohn.

Tinbergen, J. (1937): “*Econometric approach to business cycles*”. París.

Tinbergen, J. (1938): “On the theory of business-cycle control”. Econometrica, vol. 6 (22-29).

Tinbergen, J. (1939): “*Statistical testing of business-cycle theories*”. Ginebra.

Tintner, G. (1938a): “A note on economic aspects to the theory of errors in time series”.

Quarterly Journal of Economics, vol. 53 (141-149).

Tintner, G. (1938b): “The maximization of utility over time”. Econometrica, vol. 6 (154-158).

Walker, G. (1931): “On periodicity in series related terms”. Philosophiocal Transactions of the Royal Society, Serie A, vol. 131 (519-532).

Walras, L. (1874): “Economique et mécanique”

Westergaard, .L. (1890): “Grunzüge der Theorie des Statistik”. En Fisher, G. (ed.).

Wicksell, K. (1907): “Krisernas Gåta”.Statøkonomist Tedsskrift (255-286).

Wold, H. (1938): “*A study in the analysis of stationary time series*”. Almqvist-Wiksells.

Wold, H. (1948): “Oin prediction in stationary time.series”. Annals of Mathematical Statistics, vol. 19 (558-567).

Wold, H. (1949): “Statistical estimation of economic relationships”. Econometrica, vol. 17 (1-22).

Wold, H. (1951): “Dynamic systems of the recursive type: Economic and statistical aspects”. Sankhyá, vol. 11 (205-216).

Wold, H. (1954): “Causality and econometrics”. Econometrica, vol. 22 (162-177).

Yule, G.U. (1921): “On the time-correlation problems, with special reference to the cvariate-difference correlation method”. JRSS, vol. 84 (495-526).

Yule, G. U. (1926): “Why do we sometimes get nonsense correlations between time-series? A study in sampling and the nature of time-series”. JRSS, vol. 89 (1-64). Reeditado e Stuart; Kendall (1971).

Yule, G. U. (1927): “On a method investigating periodicities in disturbed series, with special application to Wolfert’s sun spot numbers”. Philosophical Transactions of the Royal Society, Serie A, vol. 226 (267-298).