DESIGUALDAD, DIVERSIDAD Y CONVERGENCIA: (MAS) INSTRUMENTOS DE MEDIDA

- MODELOS DE REGRESIÓN -*

Francisco J. Goerlich Gisbert

Correspondencia a: Francisco J. Goerlich Gisbert
Departamento de Análisis Económico e I.V.I.E.
Universidad de Valencia
Campus de los Naranjos
Av. de los Naranjos s/n (Edificio Departamental Oriental)
46022 Valencia

Tel.: 96 382 82 46
Fax: 96 382 82 49
e-mail: Francisco.J.Goerlich@uv.es
Web: http://www.uv.es/~goerlich

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.
Primera Edición Septiembre 2001
Depósito Legal: V-3928-2001

* Este trabajo recoge parte de los aspectos instrumentales de un informe más amplio titulado "Dinámica de la distribución provincial de la renta. II: La forma externa de la distribución -Evolución histórica-" realizado para el Instituto Valenciano de Investigaciones Económicas (I.V.I.E). Se agradece la financiación recibida de la DGICYT, proyecto SEC98-0895, y del Instituto Valenciano de Investigaciones Económicas.

RESUMEN

Este trabajo es complementario de Goerlich (2000a) y ambos son continuación de Goerlich (1998), si bien es autocontenido y puede ser leído de forma independiente; en él se continúa la exposición de un conjunto amplio de instrumentos con el ánimo de proporcionar un marco de referencia para una mejor comprensión de la evolución dinámica de determinadas variables económicas. El trabajo comienza describiendo un modelo de referencia básico para caracterizar cualquier variable que se mueva en dos direcciones, el análisis de varianza, para continuar examinando el concepto de convergencia-β y su puesta en práctica en el contexto de modelos de regresión.

PALABRAS CLAVE: Desigualdad, diversidad y convergencia. ANOVA. Efectos fijos. Paneles dinámicos.

ABSTRACT

This work is a complement of Goerlich (2000a) and both are a follow-up of Goerlich (1998), it is however self-contained and can be read independently. It continues offering a wide range of instruments with the aim of characterizing the dynamic evolution of an economic variable that varies in two dimensions, a cross-sectional an a temporal dimension. We begin by describing a simple reference model that acomodates such a variable, the analysis of variance, and continues by considering the concept of β-convergence and its implementation in the context of regresion models.

KEY WORDS: Inequality, divergence and convergence. ANOVA. Fixed effects. Dynamic panels.

INDICE

1 Introducción y nomenclatura.

Este trabajo es complementario de Goerlich (2000a) y ambos son continuación de Goerlich (1998), si bien es autocontenido y puede ser leído de forma independiente; en él se continúa la exposición de un conjunto amplio de instrumentos con el ánimo de proporcionar un marco de referencia para una mejor comprensión de la evolución dinámica de determinadas variables económicas. Al igual que en los trabajos mencionados el análisis se realiza a partir de la exposición de una serie de técnicas con diversos grados de sofisticación comenzando con un análisis estadístico elemental.

Aunque tomaremos como punto de referencia una variable clave en el proceso de crecimiento económico, como es la **renta *per capita***, los instrumentos que expondremos a continuación son aplicables con generalidad cuando dispongamos de una variable que se mueva en dos dimensiones. Si bien en Goerlich (1998) el análisis se realizó de forma exclusiva a partir de la utilización de conceptos tomados de la literatura de la desigualdad, que ha concentrado gran parte de sus esfuerzos en la elaboración de índices que posean determinadas propiedades (Atkinson (1970), Sen (1973), Chakravarty (1990), Cowell (1995)), este **tercer trabajo**, junto con Goerlich (2000a), toma prestados conceptos de la literatura aplicada sobre convergencia económica y busca básicamente instrumentos que nos permitan caracterizar la distribución *cross-section* de la renta *per capita* para un conjunto de individuos o unidades geográficas, tales como países o regiones, que engloben a varios individuos. Por tanto sea x la renta *per capita* objeto de estudio la finalidad es caracterizar $\phi(x)$, siendo $\phi(\bullet)$ una medida de la función *cross-section* de densidad de probabilidad de x. Hay dos características interesantes susceptibles de estudio en la evolución temporal de $\phi(x)$: (*i*) la forma cambiante en el tiempo de dicha función, y (*ii*) la dinámica intra-distribucional, es decir como una parte dada de la distribución en t transita a otra parte de dicha distribución en $t + j$. Las dos características sobre las que incidiremos son pues “forma externa” y “movilidad”. El presente trabajo y su complementario (Goerlich (2000a)) se centran básicamente en el estudio de la evolución dinámica de la forma externa de la distribución (*the external shape of the distribution*), aquí se examinarán diversas formas de caracterización de $\phi(x)$ en el contexto de **modelos de regresión**, tanto **estáticos** como **dinámicos**, centrándose fundamentalmente en el concepto de β-convergencia, mientras que en Goerlich (2000a) se examinaban básicamente estadísticos útiles para caracterizar $\phi(x)$, con especial hincapié en el concepto de σ-convergencia, así como los métodos que nos permitían inferir la forma de dicha función (δ-convergencia). El estudio de lo que sucede dentro de la distribución, es decir la movilidad, se abordará posteriormente.

Dos corrientes de literatura que han permanecido separadas, pero que hasta cierto punto son complementarias y cuyas técnicas de análisis pueden combinarse adecuadamente son: (1) la literatura tradicional sobre la desigualdad (Atkinson (1970), Sen (1973), Shorrocks (1980, 1982, 1984), Chakravarty (1990), Esteban y Ray (1993, 1994), Cowell (1995)), centrada fundamentalmente en el estudio de la distribución personal de la renta, y (2) la reciente literatura sobre la convergencia económica (Barro (1991), Barro y Sala-i-Martín (1991, 1992, 1995), Quah (1993a,b), Sala-i-Martin (1994)), preocupada por la convergencia o divergencia de la renta *per capita* o productividad de diversas unidades geográficas, ya sean regiones o países. Aunque ambas literaturas han tendido a permanecer separadas es evidente que tienen importantes puntos de contacto. Basta para ello ojear los trabajos de Esteban y Ray (1993) o Esteban (1996) sobre la polarización o los de Baumol (1986), DeLong (1988) o Quah (1996a,b, 1997) sobre la existencia de clubs de convergencia para darse cuenta de que, a grandes rasgos, se está hablando de conceptos similares, grupos de individuos o regiones que presentan peculiaridades distintas del resto. Así pues aunque la literatura sobre la desigualdad parte del individuo y la del crecimiento de una unidad espacial considerablemente más amplia, las dos tratan de estudiar la evolución en el tiempo de la distribución de una variable económica considerada de especial relevancia desde el punto de vista del bienestar o de la actividad económica. Debe ser obvio entonces que las técnicas de análisis en un tipo de literatura pueden utilizarse satisfactoriamente en el otro. De hecho algunos autores (Rabadán y Salas (1996)) han propuesto medir directamente la convergencia mediante índices de desigualdad; este enfoque, llevado hasta su extremo, podría sufrir de algunas de las críticas de Quah (1993a,b) y Esteban (1996), ya que como veremos no parece adecuado reducir el concepto de convergencia a unos pocos estadísticos.

Si bien en Goerlich (1998) se examinaron conceptos procedentes de la literatura de la desigualdad, este trabajo y su complementario (Goerlich (2000a)) utilizan fundamentalmente técnicas de análisis de la literatura aplicada sobre convergencia económica con la finalidad de examinar si la distribución de corte transversal de la renta *per capita* tiende en el tiempo hacia la igualdad en dicha renta o hacia una distribución estacionaria, así como la forma de dicha distribución. El trabajo se centra en aspectos metodológicos y prácticos, no se ofrecen aplicaciones, muy numerosas por otra parte (Barro y Sala-i-Matín (1991, 1992, 1995), Mankiw, Romer y Weil (1992)), si bien cuando requiramos de algún ejemplo este utilizará los datos de la renta *per capita* provincial de la Base de Conocimiento Económico Regional, **Sophinet**, de la Fundación BBV[2].

Antes de proseguir **dos comentarios** respecto al **contenido del trabajo** son relevantes. En **primer lugar**, si bien es cierto que, al igual que en anteriores trabajos, nuestra unidad de referencia no es necesariamente el individuo, sino la renta *per capita* de áreas geográficas que engloban a varios individuos, tales como países o regiones, sería relevante introducir la dimensión poblacional en el análisis, tal como hicimos en Goerlich (1998, 2000a), especialmente en este último trabajo donde al tema de las ponderaciones se le prestó gran atención, sin embargo esta complicación está ausente aquí y ello por diversas razones; (*i*) primero por razones meramente técnicas, la cuestión de la utilización de las ponderaciones en modelos de regresión con datos de encuesta y la inferencia estadística asociada a dichos modelos es notablemente compleja (DuMouchel y Duncan (1983), Cosslett (1993), Imbens y Lancaster (1996), Deaton (1997, Cap.-2), Wooldridge (1999, 2001)), además existen casos, con este tipo de datos, en los que es óptimo no considerar las ponderaciones (DuMouchel y Duncan (1983), Deaton (1997, Cap.-2), Wooldridge (1999, 2001)), (*ii*) segundo nuestra muestra de referencia, la renta *per capita* de regiones que engloban a varios individuos, no procede de ninguna encuesta, no ha sido obtenida mediante ningún proceso de muestreo, y la cuestión de las ponderaciones no parece haber sido estudiada por la literatura econométrica relevante, si bien las características de nuestras observaciones podrían acomodarse al enfoque de ciertos trabajos (Magee, Robb y Burbidge (1998)), y (*iii*) tercero, en el contexto de modelos de regresión con datos de regiones o países se

---

[2] Cuya dirección electrónica es http://bancoreg.fbbv.es/. Los datos de población proceden del *Anuario Estadístico* del INE.

suelen considerar distintas variables explicativas de características muy diversas, con lo que las ponderaciones que pueden parecer aceptables para una variable pueden no serlo para otra, además en modelos dinámicos se producen discontinuidades en las ponderaciones, variables en el tiempo, que no son en modo alguno obvias de resolver. Por todas estas razones **la dimensión poblacional del análisis está ausente en este trabajo**, en concordancia con la reciente literatura sobre la convergencia económica.

En **segundo lugar**, el trabajo no representa, en modo alguno, una recopilación de técnicas econométricas para el tratamiento de datos de panel, existen excelentes manuales al respecto (Hsiao (1986), Mátyás y Sevestre (1992), Balgati (1995)). Nuestro **objetivo** consiste simplemente en **caracterizar** $\boldsymbol{\phi(x)}$ en el contexto de **modelos de regresión**, tanto **estáticos** como **dinámicos**, centrándonos fundamentalmente en el concepto de **$\boldsymbol{\beta}$-convergencia**; dado que nuestra variable, x, se mueve en dos direcciones ello nos llevará a la consideración de muchas de las técnicas que se utilizan en el tratamiento habitual de datos de panel, pero sólo serán consideradas en la medida que sirvan para nuestros propósitos. No hay por tanto ninguna intención de cubrir este tipo de técnicas con generalidad.

Nomenclatura

Nuestro conjunto de observaciones de referencia se mueve en dos direcciones, el ámbito espacial y el ámbito temporal, y constituye lo que la literatura reciente (Quah (1990)) ha dado en llamar un **campo de datos** (*data field*) en el que tanto n, el número de grupos o individuos, como T, el número de periodos, son razonablemente grandes o al menos de una de una dimensión similar. En Goerlich (1998, 2000a) el análisis estadístico era realizado para cada *cross-section*, de forma que la dimensión temporal era estudiada simplemente observando la evolución en el tiempo de los estadísticos calculados para cada corte tranversal, en consecuencia no considerábamos implicaciones dinámicas, lo que nos permitía omitir el subíndice t de nuestra variable objeto de análisis. Ahora consideraremos ambas dimensiones, la espacial y la temporal, de forma simultánea, por lo que deberemos indexar nuestra variable respecto a ambas dimensiones.

Así pues supongamos que disponemos de n agrupaciones de individuos para un determinado periodo temporal, $t = 1,...,T$, cuya **renta *per capita*** designamos por $\boldsymbol{x_{it}}$, $x_{it} = Y_{it}/N_{it}$,[3] siendo Y_{it} la renta y N_{it} la población de la agrupación $i = 1,2,...,n$ en el periodo $t = 1,...,T$. Sea además $\boldsymbol{p_{it}}$ la **frecuencia relativa**, esto es, el porcentaje de población por agrupación para un año dado, $p_{it} = N_{it}/N_t$, $N_t = \Sigma_{i=1}^{n} N_{it}$, entonces la **renta *per capita* media** para el agregado en un año determinado puede expresarse como una media aritmética ponderada,

$$\mu_t = \frac{Y_t}{N_t} = \frac{\Sigma_{i=1}^{n} Y_{it}}{N_t} = \Sigma_{i=1}^{n} \frac{Y_{it}/N_{it}}{N_t/N_{it}} = \Sigma_{i=1}^{n} p_{it} x_{it} \tag{1}$$

Nuestra **variable** de referencia es por tanto la **renta *per capita***, x_{it}, de forma que realizaremos la exposición en términos de esta variable y ocasionalmente sus **pesos asociados**, p_{it}. En ocasiones nos será útil abstraernos del nivel de nuestra variable, y considerarla en términos relativos respecto al valor agregado de la misma en un año determinado, entonces deberemos normalizar x_{it} por su valor medio en el año t, μ_t, de forma que en la práctica ocasionalmente estaremos interesados en la variable $z_{it} = \frac{x_{it}}{\mu_t}$; esta es la normalización adoptada por los índices de desigualdad relativos (Goerlich (1998)). No obstante esta normalización no nos será ahora tan útil como en ocasiones anteriores.

Finalmente **dos breves reflexiones**, en primer lugar palabras como desigualdad, diversidad, diferenciación y convergencia son utilizadas como sinónimos en muchas partes del trabajo, lo que constituye un cierto abuso del lenguaje. Si la diversidad, o alternativamente la convergencia, es buena o mala, si debe aumentarse o disminuirse mediante políticas adecuadas, es algo que depende de juicios de valor y sobre lo que no nos pronunciaremos.

---

[3] x_{it} es la **renta real equivalente *per capita***, es decir ha sido adecuadamente deflactada y ajustada por las diferentes necesidades de las agrupaciones, familias o individuos. (Deaton y Muellbauer (1980)).

En segundo lugar la desigualdad y el crecimiento de las economías es un fenómeno complejo y multidimensional. Por ello, todo intento de resumir el proceso de convergencia en un único estadístico está abocado al fracaso. Quah (1993a,b) ha enfatizado satisfactoriamente este punto y a propuesto una serie de instrumentos metodológicos complementarios para analizar la evolución dinámica de distribuciones en el corte transversal (*model of explicit distribution dynamics*), parte de estos instrumentos, junto con otros muchos, son presentados en este trabajo y su complementario (Goerlich (2000a)). El trabajo se estructura en dos grandes secciones, la **sección 2** presenta un modelo de referencia básico para caracterizar cualquier variable que se mueva en dos direcciones y la **sección 3** examina el concepto de convergencia-β y su puesta en práctica en el contexto de modelos de regresión.

2 Un modelo de referencia: Efectos fijos (ANOVA).

Puesto que nuestro objetivo consiste en analizar la evolución de una variable, x_{it}, ya sea en niveles, diferencias o tasas de variación, que se mueve en dos direcciones, el ámbito espacial y el ámbito temporal, parece natural preguntarse cual es nuestro punto de partida, es decir es posible preguntarse cual es la contribución de cada una de estas dos dimensiones a la varianza de nuestra variable; de esta forma podremos examinar cuestiones tales como si los cambios observados a lo largo del tiempo son grandes o pequeños en relación a las diferencias observadas entre regiones y ello antes de introducir ningún tipo de variable explicativa en el análisis. Para ello comenzaremos nuestro trabajo con un **análisis de varianza** formulado en términos de un modelo de regresión.

Cuando disponemos de una variable que se mueve en una sola dirección, x_i, es bien conocido que el modelo

$$x_i = \alpha + u_i \qquad (2)$$

estimado por mínimos cuadrados ordinarios genera una estimación de α igual a la media muestral simple de x_i, $\hat{\alpha} = \bar{x} = \frac{\Sigma_{i=1}^{n} x_i}{n}$,[4] y un $R^2 = 0$ (Dougherty (1992)), de forma que como era de esperar la capacidad explicativa de una constante es nula. Cuando la variable de interés se mueve en dos direcciones, x_{it}, la situación es bien distinta ya que sin saber nada de la relación de x_{it} con otras variables podemos aspirar a "explicar" algo acerca de esta variable. Consideremos la extensión natural del modelo (2)

$$x_{it} = \alpha + \lambda_i + \eta_t + u_{it} \qquad (3)$$

[4] Mínimos cuadrados ponderados con ponderación $\sqrt{p_i}$ generaría como estimación de α la media muestral ponderada de x_i, $\tilde{\alpha} = \mu = \Sigma_{i=1}^{n} p_i x_i$.

donde x_{it} representa nuestra variable de interés, la renta *per capita* de la región $i = 1,2,...,n$ en el periodo $t = 1,...,T$, o cualquier otra variable que se mueva en dos dimensiones, el término λ_i representa el **efecto individual**, una **constante específica para cada individuo o región**, el término η_t representa el **efecto temporal**, una **constante específica para cada periodo de tiempo**, α es una **constante** que capta el valor medio de la variable x_{it} y u_{it} es un **componente idiosincrásico** del individuo i en el periodo t y que como primera aproximación podemos suponer que se trata de una variable aleatoria inobservable con esperanza cero e independiente e idénticamente distribuida, tanto en el corte transversal como en la dimensión temporal, no obstante por el momento consideraremos el modelo (3) más como un instrumento descriptivo que como un instrumento destinado a la inferencia estadística, por lo que las propiedades de u_{it} no serán por ahora de especial relevancia[5].

Este modelo constituye lo que en estadística se conoce como **análisis de varianza** (Seber (1977) Cap.-9) y en la terminología de la literatura econométrica sobre datos de panel como un **modelo de efectos fijos** (Hsiao (1986)). Aunque los términos λ_i y η_t son considerados como **efectos fijos** que representan **peculiaridades particulares (heterogeneidad) inobservables en una u otra dimensión** en la que se mueven los datos podrían alternativamente haber sido considerados como variables aleatorias, en cuyo caso tendríamos lo que se conoce en la literatura econométrica sobre datos de panel como un **modelo de efectos aleatorios** (Hsiao (1986)), en este caso los términos λ_i y η_t también representan peculiaridades particulares inobservables en una u otra dimensión, pero ahora afectan a la varianza de x_{it} en lugar de a su media, como sucede en el caso de que los efectos sean considerados como fijos. Tal y como indica Balestra (1992b, p.-45) **efectos fijos y aleatorios son dos formas alternativas de considerar la heterogeneidad inobservable y no pueden ser combinadas**. En este trabajo términos tales como λ_i y η_t serán siempre considerados como fijos y por tanto recogerán efectos de nivel, ya que para nuestra muestra de referencia los

---

[5] Si nuestra variable de interés se moviera en tres direcciones, x_{ijt}, introduciendo por ejemplo el ámbito sectorial, entonces el modelo (3) podría ser extendido de forma natural (Stockman (1988), Costello (1993), Marimon y Zilibotti (1996) y García-Milá y Marimon (1996)), si bien el análisis se complicaría enormemente más allá de los casos simples considerados en esta sección.

efectos fijos son siempre más fácilmente interpretables que los aleatorios (Balestra (1992a)) y además los métodos de estimación disponibles son más transparentes en este caso.

Tal y como está definido el modelo **los parámetros de (3) no están identificados**, por lo que sin restricciones adicionales esta ecuación no es estimable. Aunque lo habitual es eliminar un λ_i y un η_t que constituirán de esta forma la categoría de referencia encontramos mucho más instructivo introducir como **restricciones de identificación** $\Sigma_{i=1}^{n}\lambda_i = 0$ y $\Sigma_{t=1}^{T}\eta_t = 0$, con lo que el efecto individual λ_i representa la desviación del individuo *i* respecto a una media común dada por α y el efecto temporal η_t representa la desviación del periodo *t* respecto a dicha media común[6]. Estas restricciones no afectan a la bondad del ajuste del modelo (R^2), ni tampoco a la significación conjunta de λ_i o η_t, aunque permiten ortogonalizar los regresores en (3) y afectan a la interpretación de los coeficientes (Suits (1984), Kennedy (1986), Green y Seaks (1991)).

La **intuición** detrás de la formulación de (3) es bastante simple. Un valor de x_{it} por encima de la media para la región *i* en el periodo *t* puede ser explicado, bien por un factor individual específico de la propia región *i*, λ_i, que se supone le afecta por igual a lo largo de todo el periodo; bien por un factor agregado específico para el periodo *t*, η_t, que se supone afecta por igual a todas las regiones; o bien por un factor idiosincrásico que no es identificado por el modelo. Obviamente nuestros métodos sólo tienen capacidad explicativa si los factores idiosincrásicos son relativamente pequeños en relación al resto. De esta forma el modelo (3) recoge toda la heterogeneidad inobservable posible en cualquiera de los dos ámbitos de interés, por tanto a pesar de que los factores individuales y temporales se suponen inobservables su importancia relativa puede ser estimada.

El modelo (3) constituye pues un punto de referencia, cualquier otro modelo para x_{it} debe proporcionar una capacidad explicativa superior al R^2 obtenido a partir de (3), ya que dicha ecuación ha sido formulada sin acudir a ninguna otra variable relacionada con x_{it}. Desde este punto de vista **(3) representa la capacidad explicativa de nuestra ignorancia** y parece

---

[6] Estas restricciones implican que sólo $n-1$ efectos individuales y $T-1$ efectos temporales se estiman independientemente.

razonable que cualquier análisis de x_{it} comienze examinando dicho modelo, que representa el mínimo de lo que debemos ser capaces de explicar. En ocasiones este mínimo es bastante elevado, así por ejemplo si consideramos como variable de interés la tasa de crecimiento[7] de la renta *per capita* provincial en subperiodos decenales para el periodo 1955-1995, $n = 50$ y $T = 4$, obtenemos que $R^2 = 86.73\%$ (Goerlich (2000b)), considerando que las variables explicativas son solamente variables ficticias este porcentaje representa un gran poder explicativo y constituye el mínimo a explicar por cualquier otro modelo, en este caso nuestra ignorancia explica bastante. Cualquier modelo explicativo del crecimiento en la renta *per capita* provincial deberá ser capaz de explicar al menos el 86.73% de la variabilidad en las tasas de crecimiento de la renta *per capita* a nivel provincial.

La ortogonalización de los regresores introducida en (3) por las restricciones $\Sigma_{i=1}^{n}\lambda_i = 0$ y $\Sigma_{t=1}^{T}\eta_t = 0$ nos permite además descomponer el R^2 y examinar que parte de la variabilidad explicada es debida a los efectos individuales y que parte es debida a los efectos temporales. Finalmente señalar que una **hipótesis de interés** en este contexto consiste en examinar la homogeneidad de los individuos, H_0: $\lambda_1 = \lambda_2 = ... = \lambda_n$, y/o de los periodos temporales, H_0: $\mu_1 = \mu_2 = ... = \mu_T$. Obsérvese que estas hipótesis junto con las restricciones de ortogonalización implican que H_0: $\lambda_i = 0\ \forall i$ y que H_0: $\mu_t = 0\ \forall t$. La forma adecuada de llevar a cabo este contraste depende de las propiedades de la perturbación, u_{it}, así bajo condiciones ideales podemos utilizar los estadísticos F-estándar, pero sin ser específicos acerca de dichas propiedades siempre podemos utilizar estadísticos χ^2 consistentes frente a diversas formas de autocorrelación y/o heterocedasticidad de forma desconocida (White (1980), Hansen (1982), Newey y West (1987), Andrews (1991), Hansen (1992), De Jong y Davidson (2000)).

Consideraciones espaciales: Variables invariantes en una dirección.

Cualquier unidad geográfica se extiende en dos direcciones, latitud y longitud, ellas representan el esquema de coordenadas en las que enmarcar la superficie física, el rendimiento

[7] Obtenida como tasa anual media acumulativa.

económico de cada una de estas unidades geográficas puede ser representada por una altura determinada sobre el supuesto centro de gravedad de dicha superficie, uniendo todas estas alturas obtenemos una representación de la distribución espacial de la renta *per capita*. El gráfico 1 ilustra lo que queremos decir. Alteraciones en el tiempo de esta representación nos proporcionan una visión espacial de la evolución dinámica de la distribución de la renta *per capita* regional.

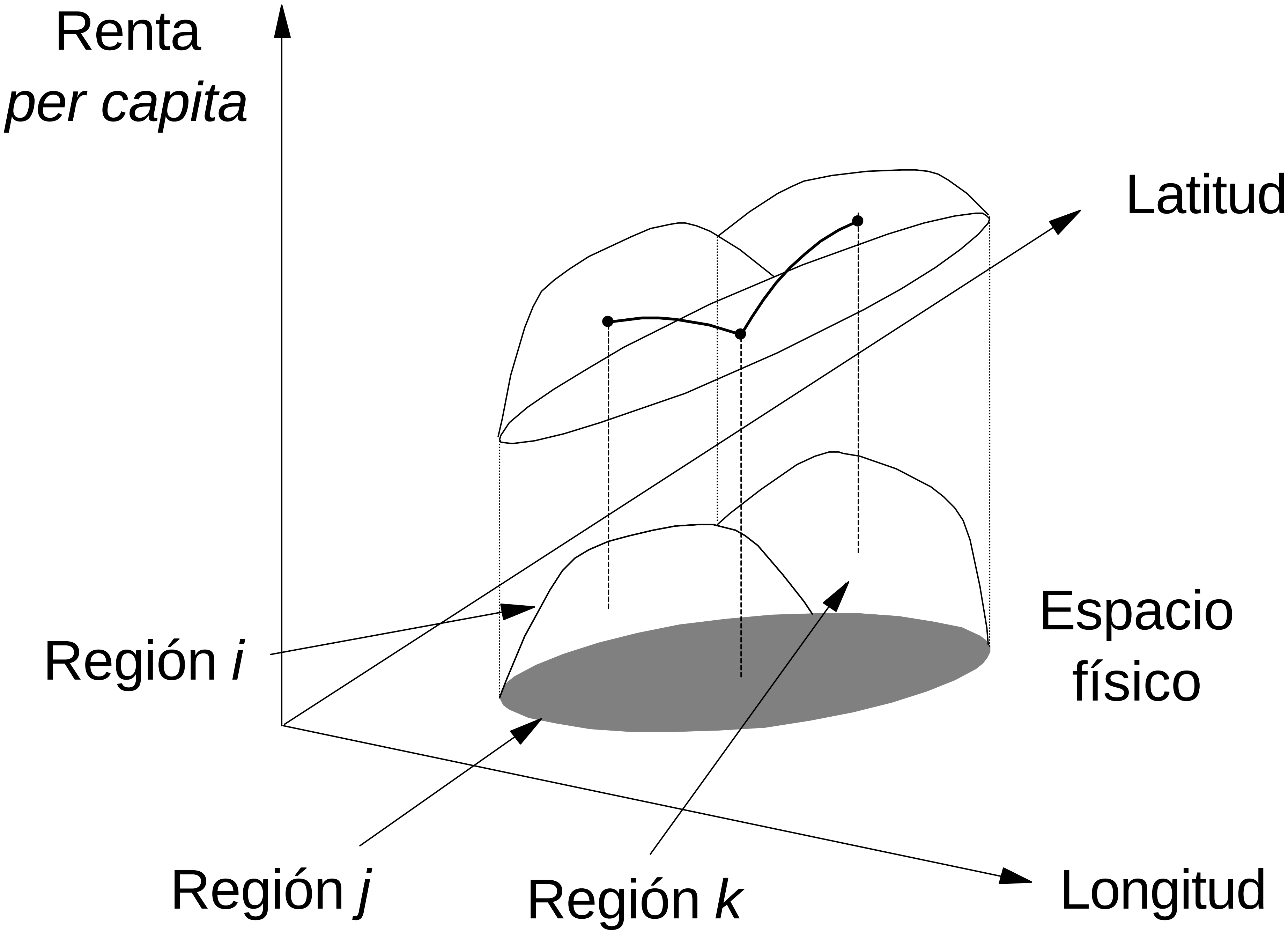
Gráfico 1

Puesto que la actividad económica tiene lugar en el espacio podríamos preguntarnos hasta que punto la superfície física que sustenta la actividad económica es un aspecto económico relevante a tener en cuenta en la determinación de las posiciones relativas entre áreas económicas. En concreto podemos estar interesados en responder a preguntas del siguiente tipo:

- ¿Juega la superficie un papel relevante en la determinación de los niveles y/o tasas de crecimiento de la renta *per capita* regional?.
- ¿Es la posición geográfica un factor clave de desarrollo?. ¿Justifica la localización el mantenimiento de diferenciales en los niveles y condiciones de vida entre regiones?.
- A pesar de que los procesos de crecimiento no son uniformes entre países o regiones vecinas es normalmente posible detectar grupos de países o regiones colindantes con niveles de renta similares en términos de su posición relativa, esto sugiere que la contigüidad puede ser importante a la hora de explicar los diferentes niveles de renta *per capita* debido a la existencia, por ejemplo, de efectos externos asociados a la actividad económica o a la mayor facilidad y menores costes en la realización de transacciones comerciales. De hecho algunos autores han sugerido que las relaciones de vecindad deben ser consideradas en los modelos económicos (Pan y LaSage (1995), Fingleton (1999c)). Ello suscita la pregunta de si las relaciones de vecindad son importantes en la determinación de los niveles de renta *per capita* de determinadas áreas económicas.

De esta forma para nuestra muestra de referencia el interés se centra en determinar como aspectos espaciales, tanto de superficie como de localización y vecindad, afectan al nivel de actividad económica, es decir a la renta *per capita* regional (Quah (1996c,d), López-Bazo,

Gráfico 1. Dinámica espacial



Vaya, Mora y Suriñach (1996), Fingleton (1999a,b)). Dicho de otra forma estamos interesados en saber si los aspectos espaciales afectan en alguna medida al nivel de renta *per capita* regional, si ello es así el papel de la política económica encaminada a reducir las disparidades en renta *per capita* deberá compensar el sesgo introducido por los factores geográficos, ya que estos nos son totalmente exógenos.

Tratar de cuantificar las preguntas anteriores equivale, en nuestro contexto y desde un punto de vista estadístico, a aumentar el modelo (3) con variables representativas de la superficie, posición geográfica o vecindad y examinar su significatividad.[8] Este procedimiento sin embargo tropieza con el problema de que este tipo de variables son invariantes en el tiempo lo que genera problemas de identificación con los efectos fijos individuales, λ_i, puesto que estas variables, digamos $\mathbf{w}_i$, siempre pueden ser escritas como combinación lineal de los λ_i. La falta de identificación puede solucionarse con la introducción de restricciones adicionales, sin embargo en el caso de las variables consideradas no hay forma de introducir dichas restricciones de forma natural, tal y como sucedía por ejemplo en el caso de los efectos fijos. Como argumenta Balestra (1992b) la introducción de variables individuales constantes en el tiempo elimina el papel jugado por las variables ficticias individuales, al menos en una forma fácilmente interpretable.

A continuación examinamos las **implicaciones de introducir** este tipo de **variables**, **invariantes en el tiempo**, en el modelo (3). Obviamente en otros contextos podemos considerar la situación simétrica de variables invariantes a través de los individuos, pero variables en el tiempo, por ejemplo en un análisis microeconómico de demanda todos los individuos se enfrentan a los mismos precios, aunque estos fluctúen a lo largo del tiempo; por razones obvias la introducción de este tipo de variables genera los mismos problemas que los que consideraremos a continuación.

Consideremos pues el modelo (3) en el que introducimos un vector de variables invariantes en el tiempo en lugar de los efectos fijos individuales, λ_i.

---

[8] Obviamente este no es el único procedimiento y a este respecto los últimos años han visto un notable desarrollo de la denominada **econometría espacial** (Florax y Rey (1995), Griffith (1996), Quah (1996d), Fingleton (1999a,b,c)) de la que no nos ocuparemos en este trabajo.

$$x_{it} = \alpha + \mathbf{w}_i'\boldsymbol{\delta} + \eta_t + u_{it} \tag{4}$$

donde $\mathbf{w}_i$ es un vector k_sx1 que contiene las variables explicativas que son constantes en el tiempo para el individuo i, excluyendo el término constante, α, y $\boldsymbol{\delta}$ es el vector k_sx1 de parámetros asociados a estas variables. A continuación se examina tanto la significatividad de estas nuevas variables introducidas, $\mathbf{w}_i$, como la bondad del ajuste del nuevo modelo. De esta forma tratamos de ver si la heterogeneidad individual inobservable que era captada por los efectos fijos individuales en (3), λ_i, es debida a estos factores o a otros no adecuadamente puestos al descubierto por estas variables. La condición de orden necesaria de identificación en (4) es que $k_s \leq n - 1$, de hecho cuando $k_s = n - 1$ el modelo (4) está exactamente identificado, en el sentido de que existe una relación uno a uno entre los efectos fijos individuales y los elementos del vector $\boldsymbol{\delta}$, por el contrario cuando $k_s > n - 1$ el vector de parámetros $\boldsymbol{\delta}$ no está identificado y no puede ser estimado (Balestra (1992b)).

Vale la pena detenernos brevemente en esta cuestión. El modelo (3) escrito en notación de muestra completa y considerando una organización de las observaciones por individuo[9], es decir primero se consideran los individuos y para cada uno de ellos se ordenan las observaciones en el tiempo, puede ser escrito como (Balestra (1992b))

$$\mathbf{x} = \ell_{nT}\alpha + \mathbf{D}_n\boldsymbol{\lambda} + \mathbf{D}_T\boldsymbol{\eta} + \mathbf{u} \tag{5}$$

donde $\mathbf{x}$ es el vector nTx1 de rentas *per capita*[10], ℓ_{nT} es un vector de unos de dimensión nTx1[11], $\mathbf{D}_n = \mathbf{I}_n \otimes \ell_T$ es una matriz nTxn que contiene el conjunto de las n variables ficticias individuales o regionales[12], $\boldsymbol{\lambda}$ es un vector nx1 de efectos fijos individuales, $\mathbf{D}_T = \ell_n \otimes \mathbf{I}_T$ es una matriz nTxT

---

[9] A menos que se indique lo contrario esta será la organización que supondremos para las observaciones.
[10] A menos que se indique los contrario nT indica organización de las observaciones por individuo mientras que Tn indica organización de las observaciones por tiempo.

[11] En general $\ell_s = \left(\overbrace{1, \ldots, 1}^{s}\right)'$.

[12] $\otimes$ representa el producto de Kronecker, de forma que

que contiene el conjunto de las T variables ficticias temporales, $\mathbf{h}$ es un vector Tx1 de efectos fijos temporales y $\mathbf{u}$ es el vector nTx1 de componentes idiosincrásicos o perturbaciones. Como ya mencionamos anteriormente estimamos (5) sujeto a las restricciones de identificación $\ell_n'\mathbf{l} = 0$ y $\ell_T'\mathbf{h} = 0$.

Por su parte el modelo (4) escrito en notación de muestra completa queda de la siguiente forma

$$\mathbf{x} = \ell_{nT}\mathbf{a} + (\mathbf{W} \otimes \ell_T)\mathbf{d} + \mathbf{D}_T\mathbf{h} + \mathbf{u} \qquad (6)$$

donde $\mathbf{W}$ es una matriz nxk_s cuya fila i-ésima es $\mathbf{w}_i'$. Utilizando las propiedades del producto de Kronecker (Magnus y Neudecker (1988), Cap.-2) observamos que

$$\mathbf{W} \otimes \ell_T = (\mathbf{I}_n \otimes \ell_T).(\mathbf{W} \otimes 1) = \mathbf{D}_n.\mathbf{W} \qquad (7)$$

de forma que podemos escribir (6) como

$$\mathbf{x} = \ell_{nT}\mathbf{a} + \mathbf{D}_n.\mathbf{W}\mathbf{d} + \mathbf{D}_T\mathbf{h} + \mathbf{u} \qquad (8)$$

Comparación entre (5) y (8) revela que la relación entre ambos modelos viene dada por

$$\left.\begin{array}{r} \mathbf{Wd = l} \\ \text{Sujeto a} \quad \ell_n'\mathbf{l} = 0 \end{array}\right\} \qquad (9)$$

$$\mathbf{D}_n = \mathbf{I}_n \otimes \ell_T = \begin{bmatrix} \ell_T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ell_T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \ell_T \end{bmatrix}$$

que es un sistema lineal de n ecuaciones en k_s incógnitas sujeto a una restricción lineal que en efecto reduce el sistema a $n-1$ ecuaciones lineales, al resultar una de ellas redundante; obsérvese que la restricción implica $\ell_n' \mathbf{W}\mathbf{d} = 0$. Por tanto la condición de orden necesaria de identificación es $k_s \leq n - 1$.

Cuando $k_s = n - 1$ el sistema (9) admite solución única y por tanto existe una relación uno a uno entre $\mathbf{l}$ y $\mathbf{d}$. Para demostrar esto simplemente introducimos la restricción $\ell_n' \mathbf{l} = 0$ en el sistema. Suponiendo, sin pérdida de generalidad, que resolvemos la restricción para el último elemento del vector $\mathbf{l}$, y particionando dicho vector como $\mathbf{l} = (\mathbf{l}^{-\prime}\, \lambda_n)'$ donde $\mathbf{l}^{-} = (\lambda_1\, \lambda_2 \ldots \lambda_{n-1})'$ de dimensión $(n-1)$x1, entonces la restricción $\ell_n' \mathbf{l} = 0$ implica que $\lambda_n = -\ell_{n-1}' \mathbf{l}^{-}$, lo que permite escribir

$$\mathbf{W}\mathbf{d} = \mathbf{l} = \begin{bmatrix} \mathbf{l}^{-} \\ \lambda_n \end{bmatrix} = \begin{bmatrix} \mathbf{l}^{-} \\ -\ell_{n-1}' \mathbf{l}^{-} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{n-1} \\ -\ell_{n-1}' \end{bmatrix} \mathbf{l}^{-} = \mathbf{E}\mathbf{l}^{-} \tag{10}$$

donde $\mathbf{E}$, de dimensión nx$(n-1)$, queda definida por la última igualdad, $\mathbf{E} = \begin{bmatrix} \mathbf{I}_{n-1} \\ -\ell_{n-1}' \end{bmatrix}$. Puesto que $\mathbf{E}'\mathbf{E} = \mathbf{I}_{n-1} + \ell_{n-1}\ell_{n-1}'$ es no singular obtenemos el sistema de $n-1$ ecuaciones como

$$(\mathbf{E}'\mathbf{E})^{-1}\mathbf{E}'\mathbf{W}\mathbf{d} = \mathbf{l}^{-} \tag{11}$$

o de forma más compacta

$$\mathbf{W}^{*}\mathbf{d} = \mathbf{l}^{-} \tag{12}$$

donde $\mathbf{W}^{*} = (\mathbf{E}'\mathbf{E})^{-1}\mathbf{E}'\mathbf{W}$, de dimensión $(n-1)$xk_s.

Cuando $k_s = n - 1$ la matriz $\mathbf{E'W}$ es cuadrada, por lo que suponiendo que $\mathbf{W}$ es de rango $n - 1$[13] dicha matriz será no singular, en este caso $\mathbf{W}^*$ puede ser invertida, $\mathbf{W}^{*-1} = (\mathbf{E'W})^{-1}\mathbf{E'E}$, con lo que obtenemos

$$\mathbf{W}^*\hat{\mathbf{d}} = \hat{\mathbf{l}}^- \qquad \Leftrightarrow \qquad \hat{\mathbf{d}} = \mathbf{W}^{*-1}\hat{\mathbf{l}}^- \tag{13}$$

para un estimador cualquiera de $\mathbf{l}^-$ y $\mathbf{d}$. En este caso ambos modelos, (3) y (4), son equivalentes y tienen la misma capacidad explicativa, por ejemplo proporcionarían el mismo R^2. Así pues cuando $k_s = n - 1$ no ganamos nada sustituyendo los efectos fijos por variables observables pero invariantes en el tiempo.

Sin embargo cuando $k_s < n - 1$ entonces el modelo (4) impone un total de $n-1-k_s$ restricciones sobre el vector $n - 1$ de efectos fijos $\mathbf{l}^-$; $\mathbf{R}\mathbf{l}^- = \mathbf{0}$, donde $\mathbf{R}$ es una matriz $(n - 1 - k_s)\text{x}(n - 1)$ de rango completo tal que $\mathbf{RW}^* = \mathbf{0}$, en otras palabras la matriz $\mathbf{R'}$ es el complemento ortogonal de $\mathbf{W}^*$, $\mathbf{R'} = \mathbf{W}^*_{\perp}$. En este caso el modelo (4) es una versión restringida de (3) y la validez de dichas restricciones puede ser contrastada mediante los procedimientos habituales. Obsérvese que siendo (4) una versión restringida del modelo (3) el R^2 en (4) no será nunca mayor que en (3) por lo que la comparación de la bondad del ajuste entre ambos modelos deberá incluir el correspondiente ajuste por grados de libertad.

[13] De otra forma algunas de las variables explicativas invariantes en el tiempo introducidas serían redundantes.

3 Implicaciones dinámicas. Convergencia-b: Una digresión.

convergencia
Acción y efecto de convergir.
converger (convergir)

1. Dirigirse dos o más lineas a unirse en un punto.
2. *fig*. Concurrir al mismo fin los dictámenes, opiniones o ideas de dos o más personas.

Diccionario de la Real Academia Española.

Si sólo dispusiéramos de dos unidades en el corte transversal entonces el análisis de la convergencia, en el sentido de la Real Academia Española, sería relativamente fácil de llevar a cabo, por el contrario si la dimensión en el corte transversal es relativamente grande, n es elevado, entonces el estudio de la convergencia se vuelve más complejo.

Si tenemos un gran número de unidades económicas en el corte transversal entonces convergencia en los niveles puede ser analizada examinando la **evolución temporal de un estadístico de dispersión** calculado para cada *cross-section*, ya que si $\mu_t > 0 \ \forall t$ entonces $CV_\omega(x) \to 0 \Rightarrow x_{it} \to \mu_t \quad \forall i$, de forma que la evolución temporal de cada x_{it} tiende a unirse en el mismo punto, μ_t, y por tanto a converger. Esta es la idea detrás del concepto de convergencia conocido en la literatura del crecimiento como **s-convergencia** (Barro y Sala-i-Martín (1995, Cap.-11.1, p.-383), Goerlich (2000a)), y que consiste por tanto en una **reducción continuada de la dispersión observada**, para una determinada variable, entre las unidades económicas objeto de análisis, lo que algunos autores han llamado "*the real test of a tendency to converge...*" (Hotelling (1933) p.-464).

Deberemos hacer, sin embargo, una **salvedad importante**, en la práctica no es posible observar $CV_\omega(x) \to 0$ sino mas bien, y en el mejor de los casos, que $CV_\omega(x) \to \alpha > 0$, es decir a cualquier efecto práctico la dispersión tiene un límite inferior positivo (Evans (1996)) y en consecuencia los niveles fluctuarán, dentro de ciertos márgenes, en torno μ_t, pero no observaremos el colapso de dichos niveles en un solo punto, la **convergencia** es ahora **a una**

distribución estacionaria, es decir invariante en el tiempo, y no a un solo punto (Quah (1993b, 1996e)), y el auténtico contraste de convergencia consiste en comprobar que la dispersión no aumenta indefinidamente (Evans (1996))[14], si bien algunos autores insisten en que la dispersión debe disminuir para hablar de convergencia (Lichtenberg (1994), Carree y Klomp (1997)).

Este concepto de convergencia, entendido como una reducción en la dispersión *cross-section* a lo largo del tiempo y conocido como **σ-convergencia** no es, sin embargo, el único concepto de convergencia al que la moderna literatura del crecimiento ha hecho referencia, otro concepto que aparece frecuentemente es el denominado **β-convergencia**[15]: diremos que existe **β-convergencia** entre un conjunto de unidades económicas, países, regiones o individuos, si existe una **relación negativa entre la tasa de crecimiento** de la renta *per capita* (o de cualquier otra variable) de dichas unidades económicas y su **valor inicial** (Sala-i-Martín (1996), p.-1327), este fenómeno es conocido también como **'regresión o reversión a la media'**. Se trata, por tanto, de un concepto esencialmente dinámico que **relaciona la situación inicial con el crecimiento posterior de una variable**.

El concepto de **β-convergencia trata de examinar si las economías inicialmente pobres**, con bajos niveles de renta *per capita* (en términos relativos respecto a un supuesto estado estacionario, o más concretamente respecto a la media del conjunto de observaciones), **han tendido a crecer más que las economías inicialmente ricas**, con altos niveles de renta *per capita*; de esta forma el concepto de β-convergencia trata de examinar si las economías pobres dan alcance (*catching-up*) a las economías ricas en términos de renta *per capita* (Barro y Sala-i-Martín (1992)).

El objeto de esta sección es múltiple: (*i*) examinar algunas implicaciones del concepto de β-convergencia, (*ii*) ver en que medida recoge la idea de *catching-up* que acabamos de

---

[14] La distinción es similar a la existente en teoría asintótica entre convergencia (*pointwise*) en probabilidad y convergencia en distribución.

[15] La terminología **σ-convergencia** y **β-convergencia** fue introducida por primera vez por Sala-i-Martín (1990) aunque los conceptos a los que hace referencia se remontan al origen mismo de la regresión (Galton (1877)). Otras acepciones del término convergencia pueden encontrarse con frecuencia en la literatura, por ejemplo en Hall, Robertson y Wickens (1992) o Quah (1993a).

mencionar, (*iii*) clarificar su relación con el concepto de σ-convergencia, (*iv*) estudiar como la β-convergencia nos ayuda a la caracterización de $\phi(x)$ y (*v*) finalmente examinar procedimientos adecuados para su estimación en el contexto de nuestro conjunto de datos. En el camino realizaremos algunas conexiones con la teoría del crecimiento, si bien **nuestra discusión se centrará en cuestiones aplicadas**. La literatura teórica del crecimiento relacionada con el concepto de β-convergencia es muy abundante (Barro (1991), Barro y Sala-i-Martin (1991, 1992, 1995), Mankiw, Romer y Weil (1992)), y existen excelentes panoramas, normalmente selectivos y sesgados, que analizan las implicaciones teóricas de la denominada β-convergencia en el contexto de los modelos de crecimiento recientes (Gould y Ruffin (1993), Andrés y Doménech (1995), Sala-i-Martín (1994, 1996), Durlauf (1996), Quah (1996e), De la Fuente (1997, 1998b), Jones (1997a,b), Jones y Manuelli (1997a,b), Pritchett (1997), Hall y Jones (1996, 1997, 1999), Durlauf y Quah (1998), Temple (1999)).

Supongamos que comparamos sólo dos momentos del tiempo, una situación inicial y una final, independientemente de que dispongamos de información sobre los periodos intermedios. En un mundo lineal y sujeto a incertidumbre podemos formalizar la idea de β-convergencia mediante la ecuación

$$g_{x_i} = \alpha - \beta x_{i,t-1} + u_{it} \tag{14}$$

donde g_{x_i} representa la tasa de crecimiento[16] de la renta per capita, $x_{i,t-1}$ la condición inicial y u_{it} un término de perturbación que captura *shocks* transitorios (estacionarios) sobre la tasa de crecimiento del individuo o región i y que como primera aproximación podemos suponer independiente e idénticamente distribuido, tanto en el corte transversal como en la dimensión temporal. La existencia de **β-convergencia implica $\beta > 0$ en (14)**, puesto que en este caso la tasa de crecimiento de x, g_x, está inversamente relacionada con la condición inicial, x_{t-1}.

Aunque no es estrictamente necesario a partir de la definición **normalmente** se restringe superiormente a β tal que **$\beta < 1$**; como se hará evidente más adelante, **$1 < \beta < 2$ implica** una

[16] Normalmente expresada en términos anuales.

situación en la que se producen **saltos periódicos** (*leapfrogging*) dentro de la distribución *cross-section*, de forma que ricos y pobres alteran sus posiciones relativas periodo a periodo hasta alcanzar el estado estacionario[17], este comportamiento no puede ocurrir en el contexto del modelo neoclásico de crecimiento que dió origen a la ecuación (14) (Barro y Sala-i-Martín (1992)), pero puede darse en los modelos de adaptación tecnológica estudiados en Barro y Sala-i-Martín (1995, Cap.-8). Por el contrario **$\beta > 2$ implica** una situación en la que de nuevo se producen **saltos periódicos** dentro de la distribución *cross-section*, pero ahora ricos y pobres alteran sus posiciones relativas periodo a periodo sin alcanzar ningún estado estacionario[18] (*overshooting*); hay β-convergencia en el sentido definido más arriba, pero no realmente convergencia, una situación de nula relevancia práctica.

Un comentario sobre el cálculo de tasas de crecimiento.

La ecuación (14) ha dejado deliberadamente sin concretar el cálculo práctico de g_x, la tasa de crecimiento, sin embargo, tal y como ha puesto de manifiesto recientemente Kakwani (1997) el cálculo de tasas de crecimiento agregadas, cuando la agregación es a través del tiempo, no es en absoluto una cuestión trivial, especialmente si deseamos realizar comparaciones en términos de bienestar. En la práctica existen numerosos procedimientos computacionales de calcular tasas de crecimiento medias para un periodo dado y muchos de esos procedimientos no sólo son capaces de alterar la magnitud del crecimiento o decrecimiento sino en ocasiones también el signo de variación.

La mayoría de los procedimientos utilizados parten de la fórmula para el cálculo de la **tasa de media de crecimiento anual acumulativo** entre dos periodos, que resuelve g_x a partir de

$$x_t = x_{t-j}(1+g_x^1)^j \qquad (15)$$

[17] En términos de series temporales el proceso estocástico para x sería (asintóticamente) **estacionario**, pero presentaría **autocorrelación negativa**.
[18] En términos de series temporales el proceso estocástico para x sería **no estacionario**, presentando además **autocorrelación negativa**.

de donde se obtiene[19],

$$g_x^1 = \left(\frac{x_t}{x_{t-j}} \right)^{1/j} - 1 = \exp\left\{ \frac{1}{j} \left(\log x_t - \log x_{t-j} \right) \right\} - 1 \qquad (16)$$

Puesto que $\log(1+g_x) \approx g_x$ para g_x en el entorno de cero, una **aproximación logarítmica** al cálculo de tasas de crecimiento cuando estas son pequeñas, frecuentemente utilizada en la práctica, viene dada por

$$g_x^2 = \frac{1}{j} \left(\log x_t - \log x_{t-j} \right) \qquad (17)$$

Sin embargo estas dos fórmulas, que son las más utilizadas, son claramente ineficientes, puesto que la tasa de crecimiento es completamente insensible a lo que ha ocurrido entre t y $t-j$, considerando de esta forma sólo la situación inicial y la final. Si deseamos introducir información sobre los años intermedios una aproximación natural consiste el calcular tasas de crecimiento periodo a periodo y obtener la **tasa media del conjunto del periodo como una media aritmética simple de las tasas de todos los periodos**, es decir

$$g_x^3 = \frac{\Sigma_{s=1}^{j} g_{x,s}^1}{j} \qquad (18)$$

o alternativamente haciendo el cambio de variable $j = t-1$ en (15), $x_t = x_1 (1+g_x)^{t-1}$, y tomando logaritmos

$$\log x_t = \log x_1 + (t-1).\log(1+g_x) \qquad (19)$$

lo que sugiere la siguiente regresión logarítmica

[19] Este es el cálculo de la función @TASA(•) de Lotus 123™.

$$\log x_t = \alpha_0 + \alpha_1 t + \varepsilon_t \qquad (20)$$

donde $\alpha_0 = \log x_1 - \log(1+g_x)$, $\alpha_1 = \log(1+g_x)$ y ε_t es un término de error. La ecuación (20) permite, una vez estimado α_1 por $\hat{\alpha}_1$, obtener una estimación de g_x como,

$$g_x^4 = e^{\hat{\alpha}_1} - 1 \qquad (21)$$

o utilizando la aproximación $\log(1+g_x) \approx g_x$ simplemente

$$g_x^5 = \hat{\alpha}_1 \qquad (22)$$ [20]

La práctica habitual estima (20) mediante **mínimos cuadrados ordinarios**, en cuyo caso la tasa de crecimiento para la totalidad del periodo es aproximadamente igual a una media ponderada de las tasas de crecimiento de los diferentes subperiodos, donde las ponderaciones son variables en el tiempo y el peso máximo se otorga a las tasas de variación del centro del periodo mientras que las menores ponderaciones se aplican al principio y al final del periodo (Kakwani (1997)), ello sugiere que otros esquemas de ponderación son posibles y quizá más deseables (Kakwani (1997)). Obsérvese que g_x también aparece en α_0 por lo que podría obtenerse igualmente la tasa de crecimiento medio a partir de una estimación de este parámetro o mediante un procedimiento que impusiera las restricciones correspondientes entre α_0 y α_1, tal como **mínimos cuadrados restringidos** (Kakwani (1997)). Adicionalmente mínimos cuadrados ordinarios puede no ser el mejor método de estimación, pudiendo examinar otros como **mínimos cuadrados perpendiculares** o variantes de **mínimos cuadrados generalizados** puesto que el término de error, ε_t, en (20) está probablemente autocorrelacionado (Canjels y Watson (1997)).

Por razones que se harán evidentes posteriormente, en esta sección utilizaremos la aproximación logarítmica g_x^2 como método de cálculo de tasas de crecimiento en la ecuación de

[20] Obsérvese que si sólo disponemos de dos observaciones en el tiempo entonces $g_x^1 = g_x^3 = g_x^4$ y $g_x^2 = g_x^5$.

β-convergencia, ya que este es el procedimiento habitual en la literatura, aunque sin lugar a dudas no es el más conveniente; de esta forma tomando $j = 1$ (14) queda formulada como

$$\log x_{it} - \log x_{i,t-1} = \alpha - \beta x_{i,t-1} + u_{it} \tag{23}$$

Linealidad en la ecuación de β-convergencia.

La ecuación que formaliza el concepto de β-convergencia, (14), postula una relación lineal entre la tasa de crecimiento y la condición inicial, sin embargo la literatura aplicada ha postulado normalmente una **relación no-lineal**, donde la **tasa de crecimiento** es **función lineal del logaritmo de la condición inicial**;

$$g_{x_i} = \alpha - \beta \log x_{i,t-1} + u_{it} \tag{24}$$

la razón estriba en la derivación de la ecuación que formaliza el concepto de β-convergencia a partir del modelo de crecimiento neoclásico y que será examinada más adelante; sin embargo desde el punto de vista meramente aplicado **esta es una cuestión empírica** a la que la literatura no ha prestado prácticamente atención y que puede resolverse simplemente examinando estadísticos de **bondad del ajuste**. En la práctica ambas ecuaciones, (14) y (24), pueden anidarse en la siguiente formulación general

$$g_{x_i} = \alpha - \beta x_{i,t-1}^{(\lambda)} + u_{it} \tag{25}$$

donde $x_{i,t-1}^{(\lambda)}$ representa la transformación de Box-Cox (1964),

$$x_{i,t-1}^{(\lambda)} = \begin{cases} \dfrac{x_{i,t-1}^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \log x_{i,t-1} & \lambda = 0 \end{cases} \tag{26}$$

En la práctica nuestra **especificación operativa** para cuantificar el concepto de β-convergencia será por tanto

$$\log x_{it} - \log x_{i,t-1} = \alpha - \beta \log x_{i,t-1} + u_{it} \qquad (27)$$

en consonancia con la literatura aplicada. Además la especificación en logaritmos permite ciertas manipulaciones algebraicas sencillas y convenientes, y para las que es posible obtener resultados exactos. Sin embargo la cuestión central es que esta especificación puede ser contrastada en un marco mucho más general y no debe tomarse como dada.

β-convergencia: **Cross-section versus *series temporales.***

Observamos que (27) puede escribirse como

$$\log x_{it} = \alpha + \rho \log x_{i,t-1} + u_{it} \qquad \rho = 1 - \beta \qquad (28)$$

un proceso AR(1) en logaritmos para cada i; por lo tanto **β-convergencia, β > 0, implica ρ < 1** en (28), más concretamente si nos restringimos a la situación habitual (Sala-i-Martín (1996)),

$$0 < \beta < 1 \quad \Leftrightarrow \quad 0 < \rho < 1$$

lo que indica que β-convergencia, en el sentido que lo hemos definido, implica, en términos de series temporales, un proceso estacionario con autocorrelación positiva para los logaritmos de x. Esto clarifica la relación, en términos del análisis aplicado estándar, entre los contrastes de series temporales y los contrastes *cross-section* de la hipótesis de convergencia.

Para futuras referencias vale la pena tener presente la relación entre ambos parámetros, β y ρ, que es la siguiente,

$$
\begin{aligned}
\beta \leq 0 &\quad\Leftrightarrow\quad 1 \leq \rho \\
0 < \beta < 1 &\quad\Leftrightarrow\quad 0 < \rho < 1 \\
\beta = 1 &\quad\Leftrightarrow\quad \rho = 0 \\
1 < \beta < 2 &\quad\Leftrightarrow\quad -1 < \rho < 0 \\
\beta \geq 2 &\quad\Leftrightarrow\quad \rho \leq -1
\end{aligned}
$$

Por lo tanto en términos de un individuo, país o región, es decir **para cada *i***, el concepto de **b-convergencia queda reducido a la estabilidad del proceso estocástico (del logaritmo) de la renta *per capita***, lo que puede ser formalizado en términos de un **contraste de raíces unidad en series temporales**, dicho con otras palabras se trata de saber si el proceso para $\log x$ tiene una distribución estacionaria o no y ello se instrumenta mediante un contraste de H_0: $\rho = 1$ ($\beta = 0$) en (28)/(27). Este es un problema que ha recibido una enorme atención por parte de la literatura macroeconómica tras la aportación inicial de Nelson y Plosser (1982) y de hecho la estacionariedad en $\log x$ es una de las implicaciones más importantes del modelo neoclásico de crecimiento económico de Solow (1956)-Swan (1956) y Cass (1965)-Koopmans (1965), desarrollado en términos de un solo país o individuo representativo[21].

Solamente **dos cuestiones prácticas**, que aparecerán posteriormente, tienen interés ahora en un contexto de series temporales:

- En la medida en que x_t presenta **crecimiento** sostenido en el tiempo este debe ser recogido bajo la hipótesis alternativa en el contraste y por tanto α esconde una función determinista del tiempo, quizá con algún proceso de ruptura (Perron (1989, 1990), Perron y Vogelsang (1992, 1993), Vogelsang (1998), Vogelsang y Perron (1998)).

- La **dinámica** en (28), un solo desfase, es probablemente demasiado restrictiva y debe ser ampliada (Evans y Karras (1996a), Evans (1997)).

---

[21] No todos los autores estarían sin embargo de acuerdo con esta afirmación (Binder y Pesaran (1996), Lee, Pesaran y Smith (1997)).

No obstante esta implicación del concepto de β-convergencia, la convergencia de un país a su propio estado estacionario, no dice absolutamente nada acerca de si las **rentas *per capita* relativas** entre países están convergiendo unas hacia las otras, por lo que es de escasa relevancia práctica y la literatura del crecimiento económico no se ha centrado en ella, aunque si es posible encontrar algunas aplicaciones (Evans (1997)). Por el contrario esta literatura ha tendido a enfatizar más la transitoriedad o no de las **diferencias entre países** o regiones, más o menos similares, que la dinámica de un solo país; de hecho una parte importante del concepto de **β-convergencia** es el examen de si la **economías pobres han tendido a crecer más que las economías ricas** (*catching-up*), por lo que el subíndice i en (28) es importante y no puede ser eliminado sin perder una parte sustancial del análisis económico.

Es por ello que los **trabajos** centrados en los contrastes de **series temporales** de la hipótesis de convergencia han seguido fundamentalmente **dos direcciones**:

(*i*) Por una parte algunos autores han examinado la **estacionariedad de largo plazo** de la ecuación (28) donde x_i **es reinterpretada como la renta relativa de dos países o regiones y $\alpha = 0$** (los paises o regiones comparten el mismo estado estacionario)[22], por ejemplo Quah (1990), Bernard y Durlauf (1991, 1995), Durlauf (1993) o Carlino y Mills (1993). Este enfoque tiene la ventaja adicional de distinguir entre convergencia entre pares de regiones y convergencia para todas las regiones, ya que algunos grupos de regiones pueden converger aunque todas en su conjunto no lo hagan, y puede resultar de interés identificar el grupo de regiones divergentes.

Es necesario puntualizar que la **aproximación de series temporales** al problema de contrastar la hipótesis de convergencia utiliza una **definición ligeramente** diferente de la que hemos identificado con β-convergencia, puesto que estos autores piensan en la convergencia no como una relación entre la tasa de crecimiento y el valor inicial de la renta *per capita* para un periodo de **tiempo fijo**, sino como una relación acerca de la predicción de largo plazo de la renta *per capita* tomando un conjunto de condiciones iniciales dadas, en concreto para dos regiones i

[22] Más concretamente el logaritmo de la **renta relativa** debe ser un **proceso estocástico estacionario con media cero** (Bernard y Durlauf (1996)).

y *j* y sus respectivas rentas *per capita*, x_i y x_j, podemos definir la **convergencia en términos de predicción** de largo plazo, ***f*-convergencia**, como (Bernard y Durlauf (1996))[23]

$$\lim_{T\to\infty} E\left(\log x_{i,t+T} - \log x_{j,t+T} \middle| \boldsymbol{I}_t\right) = 0 \tag{29}$$

siendo $\boldsymbol{I}_t$ la información en t que representa el conjunto de condiciones iniciales que tomamos como dadas al hacer la predicción. En palabras, ***f*-convergencia** implica la **igualdad en la predicción de largo plazo**, $T\to\infty$, (del logaritmo de) **de la renta *per capita* para las dos economías consideradas**, dada la información disponible en t.

Es fácil observar que

$$\lim_{T\to\infty} E\left(\log x_{i,t+T} - \log x_{j,t+T} \middle| \boldsymbol{I}_t\right) = 0 \quad \Rightarrow \quad E\left(\log x_{i,t+T} - \log x_{j,t+T} \middle| \boldsymbol{I}_t\right) < \log x_{it} - \log x_{jt} \tag{30}$$

para algún T fijo (Bernard y Durlauf (1996), Proposición 2, p.-166), es decir

***f*-convergencia Þ b-convergencia**

para estas dos economías, cuando las tasas de crecimiento son medidas entre t y $t+T$ para un horizonte temporal finito T[24]. Además el modelo de crecimiento neoclásico, usualmente utilizado en la literatura, satisface ambos conceptos de convergencia (Bernard y Durlauf (1996), Proposición 3, p.-166), por lo que ambas definiciones son útiles ya que representan diferentes implicaciones del modelo, β-convergencia para un T fijo y *f*-convergencia para $T\to\infty$.

Obsérvese que si en (28) reinterpretamos x_i como la renta relativa de dos economías, *i* y *j*, y suponemos que estas tienen el mismo estado estacionario, de forma que fijamos $\alpha = 0$ en (28), entonces esta ecuación con $0 < \rho < 1$, implica que $E\left(\log x_{it} - \log x_{jt} \middle| \boldsymbol{I}_{t-1}\right) < \log x_{i,t-1} - \log x_{j,t-1}$. Así pues, **la distinción crítica entre *f*-**

[23] Utilizando de nuevo una especificación logarítmica.
[24] La implicación contraria, sin embargo, no es cierta.

convergencia y b-convergencia consiste en darse cuenta de que la reducción esperada en el *gap* contemporáneo en la renta *per capita* relativa de dos economías no es lo mismo que la esperanza de su desaparición en el largo plazo, es por ello que los contrastes *cross-section* imponen sobre las tasas de crecimiento de la renta *per capita* restricciones más débiles que los contrastes de series temporales, ya que estos últimos están basados en un concepto de convergencia más estricto que los contrastes *cross-section*.

La **cuestión clave** es por tanto que mientras los trabajos que adoptan el **enfoque** de **series temporales** consideran un **horizonte infinito**, $T \rightarrow \infty$, los trabajos que adoptan el **enfoque *cross-section*** consideran un **horizonte fijo** entre dos momentos del tiempo. Es esta distinción la que hace que en la práctica ambos tipos de contrastes hayan tendido a generar resultados contradictorios, así mientras el **enfoque de series temporales ha mostrado la ausencia de convergencia para diversos conjuntos de datos** (Quah (1990), Bernard (1992) o Bernard y Durlauf (1995)), el **enfoque *cross-section* ha tendido a aceptar la hipótesis de convergencia tanto a lo largo del tiempo como a través de diversas muestras**, con una estimación de β bastante estable y que implica una velocidad de convergencia en torno al 2% anual (Barro y Sala-i-Martín (1992, 1995), Mankiw, Romer y Weil (1992), Sala-i-Martín (1994, 1996))[25]. De hecho, como han observado Bernard y Durlauf (1996), los contrastes de series temporales al requerir que (los logaritmos de) las rentas relativas sean un proceso estacionario con media cero necesitan una condición inconsistente con la requerida por las regresiones *cross-section*, esto es, que las diferencias entre ricos y pobres tengan una media

[25] El hecho de que velocidades de convergencia relativamente bajas y situadas en el entorno del 2% hayan aparecido en la práctica en varios trabajos aplicados utilizando diversas muestras, ya sea regionales o de países, y periodos temporales diversos (Barro (1991), Barro y Sala-i-Martín (1991, 1992, 1995), Sala-i-Martín (1994, 1996)) ha hecho que algunos autores argumenten que esta regularidad empírica puede deberse precisamente a un simple **sesgo estadístico** debido a la presencia de raíces unidad, e indique en la práctica ausencia de convergencia. De esta forma **la estabilidad de la velocidad de convergencia estaría reflejando simplemente la ausencia de la misma**. Esta explicación ha sido explorada convincentemente por Quah (1996e) y el argumento se basa en que sabemos, a partir de la teoría de la regresión en presencia de raíces unidad, que el estimador de mínimos cuadrados ordinarios de un proceso autoregresivo con una raíz unidad es consistente (en realidad “super-consistente”), pero sesgado a la baja para un T finito (Fuller (1976, Cap.-8)), lo que encaja perfectamente con un valor de ρ observado de 0.98, y en consecuencia con una velocidad de convergencia en el entorno del 2%, frecuentemente encontrada en la práctica; sin embargo dicho valor no sería más que la estimación sesgada de un valor poblacional igual a la unidad.

Otros autores han argumentado que este **sesgo estadístico** puede estar causado por la **transformación logarítmica-nolineal** usualmente utilizada en las ecuaciones de convergencia (Quah (1996e), nota 3, p.-1356).

diferente de cero; por ello es de esperar resultados contradictorios con ambos tipos de contrastes.

Así pues los **contrastes de series temporales requieren que las economías bajo estudio estén cerca de su equilibrio de largo plazo**, puesto que los contrastes suponen implícitamente que los momentos muestrales estiman adecuadamente los momentos poblacionales, **y esta es una condición incompatible con economías en transición hacia el estado estacionario**, cuando los datos están afectados en gran medida por dinámica transitoria, que es precisamente lo que suponen los contrastes *cross-section*. Esto no es sino otra forma de ver el conocido resultado de que el proceso estocástico AR(1) que representa (28) con $0 < \rho < 1$ no es ni estacionario ni asintóticamente independiente a no ser que supongamos que el proceso ha estado funcionando desde un tiempo infinitamente remoto (Spanos (1986), p.-150), de forma que cualquier dependencia respecto a las condiciones iniciales haya desaparecido. Por tanto **las aproximaciones de series temporales y *cross-section* al contraste de la convergencia descansan sobre diferentes interpretaciones de los mismos datos bajo estudio y ambos contrastes examinan el mismo coeficiente de regresión, pero visto desde diferentes perspectivas** (Leung y Quah (1996)), es por esta razón por la que tienden a generar conclusiones contradictorias.

Otros autores han transladado la cuestión de la convergencia y la existencia de raíces unidad desde los niveles en renta *per capita* a las tasas de crecimiento (Jones (1995)) o a la dispersión *cross-section* para un conjunto amplio de observaciones (Evans (1996))[26].

(*ii*) Por otra parte, en un intento de aumentar la potencia de los contrastes de raíces unidad y aprovechar la estructura de panel con una dimensión temporal razonablemente larga (*data field*) de muchos de los conjuntos de datos utilizados habitualmente en el análisis aplicado, algunos autores han tratado de contrastar el concepto de **β-convergencia** mediante la **estabilidad del proceso estocástico (del logaritmo) de la renta *per capita* para muchos países o regiones simultáneamente**, de forma que ahora la convergencia ya no es en términos

[26] Raíces unidad en σ-convergencia.

relativos de dos economías, sino por término medio para un gran conjunto de ellas. Esta idea se formaliza a través de los **contrastes de raíces unidad en datos de panel**, que a partir de los desarrollos iniciales en este campo de Levin y Lin (1992, 1993), Quah (1994b) y Breitung y Meyer (1994), han sido aplicados en el contexto de la convergencia económica por Evans y Karras (1996a, b) y Gaulier, Hurlin y Jean-Piere (1999)[27].

Muchos autores considerarían la hipótesis $H_0: \rho = 1$ $(\beta = 0)$ en (28)/(27) en este contexto de datos de panel como el elemento esencial de la hipótesis de ausencia de convergencia, al menos en el contexto del modelo de crecimiento neoclásico, pero como han señalado numerosos autores dicha conclusión es infundada (Kelly (1992), den Haan (1995), Canova y Marcet (1995), Kocherlakota y Yi (1995), Leung y Quah (1996)), de forma que al igual que sucede en la macroconomía de corto plazo no está muy claro que conclusiones teóricas podemos derivar de la existencia de raíces unidad (Quah (1987), DeLong y Summers (1988)).

Al igual que sucede con la literatura estadística acerca de los contrastes de raíces unidad en series temporales (Diebold y Nerlove (1990), Campbell y Perron (1991), McCallum (1993), Ogaki (1993)) existen ya en la actualidad un gran conjunto de estadísticos potencialmente utilizables para contrastar raíces unidad en el contexto de datos de panel (Banerjee (1999)). Estos estadísticos tienen, en lineas generales, las mismas ventajas e inconvenientes que sus homólogos en el caso de series temporales, en cuanto a escasa potencia frente alternativas locales y dependencia respecto a los componentes deterministas del proceso generador de datos, debiendo añadir además los problemas derivados de la posible heterogeneidad en el corte transversal. Todo ello hace, como es bien sabido, que diferentes estadísticos generen resultados contradictorios y en consecuencia se deba ser muy cuidadoso en el tratamiento adecuado de los

[27] La literatura teórica sobre raíces unidad y cointegración en datos de panel constituye en la actualidad un áera reciente en rápida expansión (Im, Pesaran y Shin (1997), Entorf (1997), McCoskey y Kao (1998), Benerjee (1999), Hall, Lazarova y Urga (1999), Maddala y Wu (1999), Moon y Phillips (1999, 2000), Pedroni (1999a,b), Hsiao, Pesaran y Tahmiscioglu (1999), Harris y Tzavalis (1999), Granger y Hyung (1999), Banerjee, Marcellino y Osbat (2000), Binder, Hsiao y Pesaran (2000), Karlsson y Löthgren (2000)) y cuya principal dificultad radica en la naturaleza multidimensional de la teoría asintótica que le es aplicable (Phillips y Moon (1999)). A pesar de las novedades en los desarrollos teóricos existen ya numerosas aplicaciones en campos diversos, algunas de las cuales casi han precedido a los desarrollos teóricos (Boumahdi y Thomas (1991), MacDonald (1996), Oh (1996), Culver y Papell (1997), Papell (1997), Coakley y Fuertes (1997), Pedroni (1997), Kao, Chiang y Chen (1999), McCoskey y Kao (1999), Maddala (1999), Strazicich, Co y Lee (2001)).

componentes deterministas (Campbell y Perron (1991)) y en la posible heterogeneidad entre las observaciones *cross-section* (Im, Pesaran y Shin (1997), Lee, Pesaran y Smith (1997), Maddala y Wu (1999) y Granger y Hyung (1999)).

Tal y como ilustran los trabajos de Islam (1995, 1998) y Lee, Pesaran y Smith (1997, 1998), así como la polémica entre estos autores, el tratamiento de la heterogeneidad es de vital importancia, tanto desde el punto de vista estadístico como de interpretación económica, siendo las estimaciones de la velocidad de convergencia altamente sensibles al tratamiento de la heterogeneidad.

Desde el punto de vista práctico la ecuación (28) puede ser reparametrizada convenientemente como

$$\Delta \log x_{it} = \alpha + (\rho - 1) \log x_{i,t-1} + u_{it} \tag{31}$$

donde Δ es el operador diferencia temporal, $\Delta \log x_{it} = \log x_{it} - \log x_{i,t-1}$.

Para dotar a (31) de generalidad es necesario incorporar las **dos cuestiones prácticas** que mencionamos al principio de este epígrafe y una más relacionada con la posibilidad de que las economías posean diferentes estados estacionarios. En primer lugar es necesario recoger el **crecimiento** en x_t, aunque no existe una forma única de hacerlo encontramos que la incorporación en (31) de **efectos fijos temporales** es suficientemente flexible[28], al igual que hicimos en el análisis de varianza, (3). En segundo lugar deberemos **relajar la dinámica**, que en la práctica debe ser determinada de forma empírica (Evans y Karras (1996a), Evans (1997)).

[28] También podríamos introducir **tendencias temporales heterogéneas** para cada una de las unidades *cross-section*, tal y como hacen Levin y Lin (1992) o Harris y Tzavalis (1999), y que sería lo apropiado en el caso de series temporales como mecanismo para discriminar entre series estacionarias en tendencia lineal y series estacionarias en diferencias (Phillips y Perron (1988)). Alternativamente también podríamos centrarnos en la variable z_{it} en lugar de x_{it} (Levin y Lin (1993), Evans y Karras (1996a), Gaulier, Hurlin y Jean-Pierre (1999)), de esta forma habríamos eliminado las tendencias presentes en los datos y nos centraríamos en el comportamiento de las desviaciones respecto al valor medio del agregado. En nuestro caso encontramos más flexible la introdución de efectos fijos temporales ya que permiten recoger con más facilidad tendencias no lineales y diversos procesos de ruptura estructural, que no obstante se suponen comunes a todas la unidades *cross-section* y en consecuencia no recogen la heterogeneidad en los procesos de crecimiento.

Finalmente parece restrictivo suponer, al menos inicialmente, que todas las economías tengan el mismo estado estacionario y en consecuencia es conveniente introducir en (31) **efectos fijos individuales**. Posteriormente, si $H_0: \rho = 1$ es rechazada en favor de $H_1: \rho < 1$ podremos contrastar si estos efectos fijos son significativos, lo que permitirá arrojar luz sobre la cuestión de si las economías convergen a diferentes estados estacionarios, **convergencia condicional**, o a uno solo, **convergencia absoluta o incondicional**. El epígrafe siguiente justificará, en el contexto del modelo neoclásico, la introducción de estos efectos como forma de recoger la heterogeneidad inobservable entre economías y en consecuencia como forma de distinguir entre **convergencia condicional** y **convergencia absoluta**, a pesar de las críticas que esta distinción ha suscitado en términos de la interpretación del propio concepto de β-convergencia (Durlauf y Quah (1998), Section 5).

Todas estas consideraciones sugieren ampliar (31) de la siguiente forma

$$\Delta \log x_{it} = \alpha + \lambda_i + \eta_t + (\rho - 1) \log x_{i,t-1} + \Sigma_{j=1}^{p} \theta_j \Delta \log x_{i,t-j} + u_{it} \qquad (32)$$

donde λ_i y η_t tienen la misma interpretación que en (3) y al igual que en esta ecuación introducimos $\Sigma_{i=1}^{n} \lambda_i = 0$ y $\Sigma_{t=1}^{T} \eta_t = 0$ como restricciones de identificación. Esta es la ecuación básica para contrastar raíces unidad en un contexto de datos de panel que no obstante puede ser generalizada en diversas direcciones.

Es importante observar que el contraste de $H_0: \rho = 1$ en (32) requiere que el coeficiente autoregresivo, ρ, sea homogéneo, tanto bajo la hipótesis nula como bajo la hipótesis alternativa, puesto que agrupa los datos en la dimensión *cross-section*. Im, Pesaran y Shin (1997) y Maddala y Wu (1999) consideran contrastes que no agrupan los datos en el corte transversal, sino que se basan en *n* contrastes individuales de raíces unidad, uno para cada *cross-section*. Es necesario recordar una vez más que la existencia de heterogeneidad puede distorsionar ampliamente los resultados cuando no es incorporada al análisis (Pesaran y Smith (1995), Lee, Pesaran y Smith (1995, 1997, 1998), Zietz (2001)).

Sea $\hat{\rho}-1$ el estimador de mínimos cuadrados ordinarios de $\rho-1$ en (32) y $t_{\rho=1}$ su *t-ratio* asociado, utilizado normalmente para contrastar la hipótesis H_0: $\rho = 1$. Levin y Lin (1992) han derivado las distribuciones asintóticas de $\hat{\rho}-1$ y $t_{\rho=1}$ bajo H_0: $\rho = 1$[29]. Estos autores demuestran que $t_{\rho=1}$ converge asintóticamente en distribución a una normal conforme $n, T \rightarrow \infty$ con tal que $\frac{\sqrt{n}}{T} \rightarrow 0$, por tanto se permite que la dimensión temporal se expanda más lentamente que la dimensión *cross-section*. La convergencia es relativamente rápida y se produce a la tasa $T\sqrt{n}$. La media de la distribución asintótica de $t_{\rho=1}$ está desplazada hacia valores negativos en relación a la distribución normal estándar, su varianza es inferior a la unidad y el desplazamiento en la media de la distribución asintótica es proporcional a $\sqrt{n}$. Además dicha distribución es independiente del valor de p y por tanto de los θ's, con tal que los resíduos de la versión estimada de (32) sean empíricamente blancos, así como de si los efectos fijos temporales son incluidos o no en el modelo y la forma que estos toman[30]. Harris y Tzavalis (1999) derivan las distribuciones de estos estadísticos cuando $n \rightarrow \infty$ pero la dimensión temporal, T, es fija.

Es sencillo explicar porque la porque la distribución asintótica de $t_{\rho=1}$ está desplazada hacia valores negativos en relación a la distribución normal estándar. Sólo los datos de cada región, $\log x_{i1}$, $\log x_{i2}$, ... $\log x_{iT}$, son relevantes en la estimación del efecto fijo de dicha región[31], λ_i; incluir dichos efectos fijos en (32) con $n > 1$ es enteramente análogo a incluir un término constante en (32) con $n = 1$. En este último caso Dickey y Fuller (1979) han demostrado que la distribución asintótica del estadístico *t-ratio* para contrastar la hipótesis nula de raíz unidad está desplazada hacia valores negativos, el resultado para datos de panel no es más que una extensión natural del caso de series temporales.

Si $\rho < 1$ entonces la teoría asintótica estándar es de aplicación (Hsiao (1986)). Ello implica que para contrastar la hipótesis H_0: $\lambda_1 = \lambda_2 = \lambda_3 = ... = \lambda_n$ en (32) podemos utilizar la

---

[29] Al igual que en el caso de los contrastes para series temporales esta distribución incorpora implícitamente el supuesto de que $\alpha = \lambda_i = 0 \;\; \forall i$, aunque estos efectos fijos no están restringidos bajo H_1.

[30] La distribución sería diferente si incluyeramos **tendencias temporales heterogéneas** para cada una de las unidades *cross-section* (Levin y Lin (1992), Harris y Tzavalis (1999), Maddala y Wu (1999)).

[31] Junto con la constante del modelo, α, dada la forma en que los efectos fijos están definidos.

ratio F habitual y considerarla aproximadamente distribuida como una F-Snedecor con $n-1$ grados de libertad en el numerador y $n.T-k$ en el denominador, siendo k el número de regresores en la ecuación (32)[32], alternativamente podríamos utilizar una versión del contraste robusta frente a heterocedasticidad de forma desconocida (White (1980)), que tomaríamos distribuida como una χ^2 con $n-1$ grados de libertad bajo H_0: $\lambda_1 = \lambda_2 = \lambda_3 = ... = \lambda_n$. Todo ello condicionado en $\rho < 1$.

En la práctica las distribuciones en muestras finitas pueden diferir apreciablemente de las distribuciones asintóticas (Harris y Tzavalis (1999)) por lo que es recomendable obtener los niveles de significación mediante simulación de Monte Carlo (Evans y Karras (1996a,b), Gaulier, Hurlin y Jean-Piere (1999)).

No nos extenderemos más sobre los resultados acerca de los contrastes de raíces unidad en un contexto de datos de panel y su relación con el tema de la convergencia económica, pero basten los anteriores comentarios para resaltar que el contraste de la hipótesis de ausencia de convergencia, H_0: $\beta = 0$ en (27), no es en modo alguno trivial.

b-convergencia y el modelo neoclásico de crecimiento.

La ecuación (27) no es la más usualmente utilizada en trabajo aplicado relacionado con el concepto de β-convergencia. En la práctica la **ecuación estimada** es de la forma[33]

$$\frac{1}{j}\left(\log x_{it} - \log x_{i,t-j}\right) = a - \left(\frac{1-e^{-bj}}{j}\right)\log x_{i,t-j} + \varepsilon_{it} \tag{33}$$

donde $b > 0$ asegura la estabilidad. Esta ecuación es considerada como la **implicación observable más importante del modelo de crecimiento neoclásico** (Barro y Sala-i-Martín

[32] Técnicamente la *ratio F* estándar converge a una distribución χ^2 con $n-1$ grados de libertad dividida por $n-1$, el número de restricciones a contrastar, confome $T\rightarrow\infty$ y mientras n permanece fijo en un valor finito dado; como es habitual dicha distribución puede tomarse aproximadamente distribuida como una F-Snedecor con $n-1$ grados de libertad en el numerador y $n.T-k$ grados de libertad en el denominador.
[33] Esta es la ecuación (15) en Barro y Sala-i-Martín (1992, p.-229).

(1992)), aunque obsérvese que dicha implicación hace referencia a una sola economía no a un conjunto de ellas, ya sean países o regiones.

El coeficiente de $\log x_{i,t-j}$ en la ecuación (33) es $-\left(\frac{1-e^{-bj}}{j}\right)$, que disminuye en magnitud conforme aumenta el lapso temporal, j, entre los periodos inicial, $x_{i,t-j}$, y final, x_{it}, para un $b > 0$ dado. Conforme j aumenta, el efecto de la condición inicial sobre la tasa media de crecimiento disminuye, cuando $j \to \infty$ el coeficiente $-\left(\frac{1-e^{-bj}}{j}\right) \to 0$ con tal que $b > 0$, que constituye la condición de estabilidad en (33). Debido a que la estimación de b tiene en cuenta la distancia temporal, j, entre las condiciones inicial y final, estimaciones procedentes de diferentes muestras o periodos temporales son comparables entre sí, independientemente de dicha distancia, lo que constituye una de las principales ventajas de la estimación no lineal de b a partir de la especificación (33). La otra ventaja fundamental deriva de la propia interpretación del coeficiente b como **velocidad de convergencia** hacia el estado estacionario, de esta forma si j son periodos anuales y $b = 0.02$, la velocidad de convergencia es del 2% anual, lo que quiere decir que cada año se reduce en un 2% la distacia entre el logaritmo de la renta *per capita* actual y el logaritmo de la que correspondería al estado estacionario, a esta velocidad tardaríamos algo más de 34 años en cerrar la mitad del *gap* existente entre el logaritmo de la renta *per capita* actual y el correspondiente al estado estacionario[34].

---

[34] El número de años que tardaríamos en cerrar la mitad de la distancia al estado estacionario es calculado de la siguiente manera, si x^* representa la renta per capita correspondiente al estado estacionario, la dinámica de transición implícita en la versión determinista de la ecuación (33), donde el estado estacionario se encuentra oculto en a, viene dada por la ecuación (Chiang (1984, Cap.-14), Barro y Sala-i-Martín (1992))

$$\log x_{it} - \log x^* = e^{-bj}(\log x_{i,t-j} - \log x^*)$$

el tiempo que tardaremos en recorrer la mitad del camino, $\log x_{it} - \log x^* = 0.5 \times (\log x_{i,t-j} - \log x^*)$, se obtiene resolviendo por j la ecuación anterior una vez hemos sustituido la distancia que queremos recorrer. En nuestro caso

$$0.5 \times (\log x_{i,t-j} - \log x^*) = e^{-0.02j}(\log x_{i,t-j} - \log x^*) \quad \Rightarrow \quad 0.5 = e^{-0.02j} \quad \Rightarrow \quad j = -\frac{\log 0.5}{0.02} = 34.66$$

Para futuras referencias resulta útil examinar como la ecuación (33) es derivada y la dinámica de la misma analizada, las referencias abundan (Blanchard y Fisher (1989, Cap.-2), Barro y Sala-i-Martín (1992, 1995)) por lo que sólo esgrimiremos los argumentos principales. El modelo de Solow (1956)-Swan (1956) de agentes optimizadores desarrollado por Cass (1965)-Koopmans (1965) es resuelto en tiempo continuo para una economía y la dinámica de transición analizada mediante log-linearización de una versión determinista del modelo alrededor de su estado estacionario, a partir de esta log-linearización es fácil observar que la tasa de crecimiento del *output* por trabajador, que en el contexto del modelo coincide con la renta *per capita*[35], depende de la log-desviación respecto al estado estacionario. Como es usual esta dependencia es parametrizada por el autovalor negativo de la matriz de primeras derivadas, la negatividad del autovalor asegura la estabilidad del modelo, lo que es equivalente a la condición $\beta > 0$, y la log-linearización elimina cualquier no linealidad en el modelo provocando que la convergencia al estado estacionario sea directa, sin oscilaciones ni *overshooting*, lo que es equivalente a la condición $\beta < 1$.

Una **cuestión importante** de la argumentación anterior es que el parámetro ***a* en (33) no es** en realidad **constante**, sino que **depende del estado estacionario**, y por tanto **se ve afectado por cualquier variable y/o cambio en los parámetros del modelo que afecten a dicho estado estacionario**, en concreto en un contexto temporal a presenta una tendencia en el tiempo debido a la existencia de progreso técnico y en un contexto *cross-section* a es independiente de i, $a = a_i \; \forall i$, si y sólo si todas las economías comparten el mismo estado estacionario. En general, por tanto a será una función de variables, tanto en el tiempo como en el corte transversal, de forma que $a = a(\omega_{it})$ siendo $a(\bullet)$ una función que engloba todas aquellas variables que afectan al estado estacionario de las diferentes economías, ω_{it}. Esta puntualización es la que sustenta la diferencia que aparece en la literatura entre **β-convergencia absoluta o incondicional**, cuando se supone que el estado estacionario es el mismo para todas las economías bajo estudio, $a = a_i \; \forall i$, y por tanto a es constante en ambas direcciones, salvo por la tendencia temporal que recoge el progreso técnico y que se supone en este caso común a todas las economías, y **β-convergencia condicionada**, cuando incluimos en la ecuación (33), o

[35] Ambos conceptos son sin embargo muy diferentes en las economías reales, tal y como han señalado Paci (1997) o Goerlich y Mas (1998).

alternativamente en la (27), variables explicativas adicionales que tratan de recoger diferencias en el estado estacionario de las diferentes economías. En el caso de que estas variables explicativas potenciales sean inobservables es siempre posible introducir en (33) efectos fijos individuales, cuya incorparación al análisis es normalmente recomendable para recoger cualquier tipo de heterogeneidad no observable.

Como mencionamos en la introducción el análisis de variables condicionantes y su influencia en el proceso de convergencia será tratado con posterioridad[36] (Goerlich (2001b)) por lo que en esta sección **consideraremos sólo** el concepto de **β-convergencia absoluta o incondicional**, y en consecuencia supondremos que el estado estacionario es común a todas las economías, *a* es constante, salvo por la introducción de variables ficticias que trataran de agrupar economías con características similares y que constituye en la práctica una forma particular de condicionar en la ecuación de β-convergencia, un condicionamiento basado en nuestra ignorancia.

La relación entre (27) y (33) es fácil de derivar, a partir de (28) y sustituyendo recursivamente *j* periodos hacia atrás

$$\log x_{it} = \alpha\Sigma_{k=0}^{j-1}\rho^k + \rho^j \log x_{i,t-j} + \Sigma_{k=0}^{j-1}\rho^k u_{i,t-k} \tag{34}$$

restando $\log x_{i,t-j}$ a ambas partes de la ecuación y diviendo por el lapso temporal, j,

$$\frac{1}{j}\left(\log x_{it} - \log x_{i,t-j}\right) = \frac{1}{j}\alpha\Sigma_{k=0}^{j-1}\rho^k - \left(\frac{1-\rho^j}{j}\right)\log x_{i,t-j} + \frac{1}{j}\Sigma_{k=0}^{j-1}\rho^k u_{i,t-k} \tag{35}$$

a partir de lo cual observamos que

---

[36] Obsérvese de pasada que numerosos autores han mostrado la inestabilidad y falta de robustez de las ecuaciones de convergencia condicionadas, Levine y Renelt (1992), Easterly, Kremer, Pritchett y Summers (1993), Andrés, Doménech y Molinas (1996), Temple (1998), Doppelhofer, Miller y Sala-i-Martín (2000), entre otros.

$$(i) \qquad a = \frac{1}{j}\alpha\Sigma_{k=0}^{j-1}\rho^k = \frac{1}{j}\alpha\Sigma_{k=0}^{j-1}(1-\beta)^k$$

$$(ii) \qquad \frac{1-e^{-bj}}{j} = \frac{1-\rho^j}{j} = \frac{1-(1-\beta)^j}{j} \quad \Rightarrow \quad b = -\log\rho = -\log(1-\beta)$$

$$(iii) \qquad \varepsilon_{it} = \frac{1}{j}\Sigma_{k=0}^{j-1}\rho^k u_{i,t-k} = \frac{1}{j}\Sigma_{k=0}^{j-1}(1-\beta)^k u_{i,t-k}$$

por lo que existe una relación uno-a-uno entre el parámetro β en (27) y el parámetro b en (33), que como ya hemos mencionado indica la **velocidad de convergencia** hacia el estado estacionario, de forma que conforme

$$\beta \to 0 \quad (\rho \to 1) \quad \Rightarrow \quad b \to 0$$

y

$$\beta \to 1 \quad (\rho \to 0) \quad \Rightarrow \quad b \to \infty$$

de esta forma la condición $0 < \beta < 1 \quad (0 < \rho < 1) \quad \Rightarrow \quad 0 < b < \infty$. Así pues **b = 0** ($\rho = 1$) implica **nula velocidad de convergencia**, es decir ausencia de la misma, por el contrario **b = 1** ($\rho = 0$) implica un ajuste instantáneo, la **convergencia es inmediata** y $\log x_{it}$ carece de correlación serial[37].

Puesto que (33) implica una estimación no lineal, en ocasiones se estima simplemente esta ecuación en su forma lineal

[37] El hecho de que la relación entre ρ y b sea no-lineal, y el intervalo (0,1) dentro del cual varía ρ sea transformado en el intervalo $(0,\infty)$ dentro del cual varía b, hace que muy pequeñas alteraciones en el valor de ρ sean compatibles con velocidades de convergencia bastante dispares. En concreto, con una precisión de dos decimales valores de ρ de 0.98 son compatibles con velocidades de convergencia situadas entre el 1.61% y el 2.53%.

$$\frac{1}{j}\left(\log x_{it} - \log x_{i,t-j}\right) = a - b' \log x_{i,t-j} + \varepsilon_{it} \qquad (36)^{38}$$

donde $b' = \frac{1-e^{-bj}}{j} = \frac{1-\rho^j}{j} = \frac{1-(1-\beta)^j}{j} > 0$ indica la existencia de β-convergencia. Por tanto el parámetro ρ de (28) subyacente en (36) y que captura la inercia en $\log x_{it}$ periodo a periodo viene dado por $\rho = (1 - jb')^{1/j}$.

Desde un **punto de vista** meramente **práctico** obsérvese que el paso de b' a ρ sólo tiene sentido cuando $jb' < 1$ de forma que $\rho = (1 - jb')^{1/j} > 0$; lo contrario no es cierto, el paso de ρ a b' tiene sentido aún cuando $-1 < \rho < 0$, ya que en este caso el AR(1) dado por (28) es estacionario en el largo plazo, aunque presenta autocorrelación negativa y por tanto oscilaciones, puesto que ρ^j se alterna en signo, sin embargo estas oscilaciones han sido descartadas *a priori*. Además sólo cuando $jb' < 1$, o equivalentemente $\rho > 0$, es posible obtener un valor de b interpretable en términos de la velocidad de convergencia ya que en otro caso b no está definido puesto que $b = -\log\rho$. Algunos autores que estiman directamente (36) interpretan situaciones en las que $jb' > 1$ como situaciones de "**hiperconvergencia**" (Alvarez de Toledo, Rojo, Toribio y Usabiaga (2000), p.-14), sin embargo, como veremos a continuación, esta interpretación no tiene un fundamento claro y podría interpretarse igualmente como situaciones en las que se producen saltos periódicos dentro de la distribución *cross-section*, de forma que ricos y pobres alteran su posición relativa periodo a periodo, pero no sabemos *a priori* si hay convergencia o no hacia un estado estacionario (*leapfrogging* o *overshooting*). Tal y como argumentan Leung y Quah (1996) y Quah (1996e, p.-1359) es probablemente más adecuado estimar (33) por métodos no lineales[39] que (36) por métodos lineales[40], de forma que situaciones de hiperconvergencia se muestren como valores extremadamente elevados de b.

---

[38] En ocasiones se estima

$$\frac{1}{j}\left(\log x_{it} - \log x_{i,t-j}\right) = a - b' \frac{1}{j} \log x_{i,t-j} + \varepsilon_{it}$$

[39] Tal como **mínimos cuadrado no lineales**.

[40] Por ejemplo **mínimos cuadrados ordinarios**.

La razón de la indefinición mencionada en el párrafo anterior radica en que el término e^{-bj} en (33) procede de una log-linearización en **tiempo continuo** y en este caso la estabilidad de la relación depende de que $e^{-bj} \to 0$ conforme $j \to \infty$, lo que a su vez requiere que $b > 0$, mientras que el término β en (27) o ρ en (28) es una aproximación a la dinámica de transición del modelo en **tiempo discreto** y en este caso la estabilidad de la relación depende de que $0 < \beta < 2$ ó $-1 < \rho < 1$; puesto que **las ecuaciones diferenciales de primer orden no pueden presentar comportamiento oscilatorio** pero las ecuaciones en diferencias si (Chiang (1984), Sec.-16.3), la compatibilidad entre ambos parámetros exige limitar el rango de variación de los mismos a aquellas situaciones en las que la ecuación en diferencias (28) es estable y no oscilatoria, esto es $0 < \beta < 1$ ó $0 < \rho < 1$.

Un **comentario final** que conviene tener presente en lo que hace referencia a la **interpretación teórica de la ecuación de convergencia** (33). Ya hemos observado como dicha ecuación es derivada para una sola economía, Solow (1970, p.-3) ha enfatizado este punto. La literatura aplicada, por el contrario, a estimado mayoritariamente la ecuación (33) a partir de observaciones para un conjunto dispar de unidades económicas, ya sean países o regiones, con el ánimo de arrojar luz sobre la cuestión del *catching-up*. Dicha aplicación requiere una total homogeneidad de las unidades económicas subyacentes al análisis que no es probable que se de en la práctica, ya que el supuesto implícito es que todas la unidades económicas de la muestra utilizada tienen el mismo estado estacionario (convergencia absoluta). Es posible introducir cierto grado de heterogeneidad permitiendo diferencias en el estado estacionario (convergencia condicional), de forma que $a = a(\omega_{it})$ en (33) siendo $a(\bullet)$ una función que engloba todas aquellas variables que afectan al estado estacionario, ω_{it}; pero aún así mantenemos cierto grado de homogeneidad, la velocidad de convergencia b es la misma entre economías. Permitir una completa heterogeneidad, en términos de niveles, tasas de crecimiento y velocidades de convergencia, es posible desde un punto de vista econométrico (Lee, Pesaran y Smith (1995, 1997, 1998)) pero vacía de contenido económico el propio concepto de β-convergencia en lo que hace referencia a su dimensión *cross-section* (Islam (1998)), que es de la que hemos partido al principio de esta sección. Encontrar que las economías convergen a diferentes lugares, a distintas tasas de crecimiento y con velocidades diferentes no es probablemente una conclusión muy alentadora. **Heterogeneidad e interpretabilidad del concepto de β-convergencia presentan un *trade-off* al que hay que hacer frente en cada muestra concreta**.

En consecuencia **la aplicación de la ecuación de convergencia a un conjunto de economías y el tratamiento adecuado de la heterogeneidad reflejará siempre la tensión subyacente entre la dimensión temporal, referida a cada unidad económica particular, y la dimensión *cross-section* del propio concepto de β-convergencia**, y que no es sino una forma alternativa de ilustrar las aproximaciones de series temporales y *cross-section* al concepto de β-convergencia. Aún así, deberemos recordar que el modelo de crecimiento neoclásico es un modelo para una sola economía y su aplicación a un conjunto de ellas requiere necesariamente cierto grado de homogeneidad (Islam (1998)). Al fin y a la postre quizá en el contexto del modelo neoclásico sólo tenga sentido el análisis de experiencias de crecimiento individuales (Young (1992, 1995), Hulten y Srinivasan (1999)).

Un comentario acerca de las tendencias y el progreso técnico.

Ya hemos observado en el epígrafe anterior que aún cuando supongamos que todas las economías poseen el mismo estado estacionario y por tanto $a = a_i \ \forall i$ en (33) éste parámetro recoge una tendencia temporal derivada de la existencia de progreso técnico (Barro y Sala-i-Martín (1992), p.-230)[41], y aunque dicha tendencia no es de importancia en los contrastes *cross-section* si deberá ser tenida en consideración en los contrastes de series temporales o cuando combinemos los datos en ambas direcciones. La forma en la que esta tendencia es introducida en el modelo no es, en absoluto, una cuestión trivial. Así el **tratamiento estándar** de la dinámica de transición en el modelo de crecimiento neoclásico (Barro y Sala-i-Martín (1992, 1995), King y Rebelo (1993)) postula que **el progreso técnico crece a una tasa constante y exógena** generando de esta forma la ecuación de convergencia (33) e introduciendo una **tendencia lineal determinista** en a, si añadimos incertidumbre **dicha tendencia es fruto de que el progreso**

[41] En ausencia de progreso técnico la tendencia está ausente en a y las consideraciones de este epígrafe pueden ser ignoradas.

técnico puede ser representado por un proceso estocástico estacionario en tendencia (Nelson y Plosser (1982)); por el contrario **si** suponemos que **la tecnología puede ser representada por un proceso estocástico estacionario en diferencias** (Nelson y Plosser (1982)), es decir la **tendencia** en el progreso técnico es **estocástica** o posee una **raíz unidad**, **entonces la ecuación de convergencia (33) colapsa**, **en el sentido de no ser estable** (King, Plosser y Rebelo (1988a, b), Kocherlakota y Li (1995)), a menos que (el logaritmo de) la renta *per capita* se defina en términos de eficiencia (Bernard y Durlauf (1996), Proposición 1, p.-164), lo cual no es factible en términos empíricos. En este caso (el logaritmo de) **la renta *per capita* posee una raíz unidad** (King, Plosser, Stock y Watson (1991), Campbell (1994)), y de forma similar a lo que ocurre con la relación entre consumo y renta cuando existen raíces unidad en la denominada paradoja de Deaton (1987), es de esperar que un país con un nivel de renta *per capita* más elevado que otro muestre mayores tasas de crecimiento en el futuro (Kocherlakota y Li (1995), Proposición 1, p.-213), de esta forma obtendríamos una relación positiva entre la tasa de crecimiento de la renta *per capita* y su valor inicial, es decir **b-divergencia**. El resultado contrario puede derivarse para los modelos de crecimiento endógeno, donde si los *shocks* tecnológicos son suficientemente temporales, puede darse la situación de que un país con un nivel de renta *per capita* más elevado que otro muestre menores tasas de crecimiento en el futuro (Kocherlakota y Li (1995), Proposición 2, p.-213), obteniendo de esta forma una relación negativa entre la tasa de crecimiento de la renta per capita y su valor inicial, es decir β-convergencia. Por tanto la estabilidad o no de la ecuación (23)/(33) está ligada a la persistencia o temporalidad de los *shocks* tecnológicos, existencia o no de una raíz unidad en el proceso estocástico del progreso técnico, pero no a una clase determinada de modelos de crecimiento económico.

Numerosos autores han identificado diversos problemas econométricos con las ecuaciones de β-convergencia en relación a su interpretación en términos de modelos de crecimiento económico estructurales y derivados del hecho de que **la forma en como se aumenta el modelo de crecimiento con perturbaciones estocásticas altera de forma**

sustancial las inferencias teóricas que es posible inferir de los datos (Kelly (1992), den Haan (1995), Kocherlakota y Li (1995), Leung y Quah (1996), Lee, Pesaran y Smith (1997))[42].

Por lo tanto no está excesivamente claro, especialmente si consideramos que la tecnología puede ser de carácter no estacionario, cual es la conexión entre la ecuación de convergencia (23)/(33) y el modelo de crecimiento neoclásico, ni si dicha ecuación es apropiada para discriminar entre modelos alternativos (Kocherlakota y Li (1995)). No parece pues adecuado utilizar **(33)** para realizar inferencias teóricas sobre los modelos de crecimiento, al menos sin aumentar la estructura que deberemos imponer sobre los datos, sino simplemente como un **estadístico descriptivo** más para un T fijo.

Los contrastes de convergencia basados en series temporales del tipo de los realizados por Bernard y Durlauf (1991, 1995) o Carlino y Mills (1993), al concentrarse en la renta *per capita* relativa de pares de países o regiones, se acomodan mejor a situaciones no estacionarias, al hacer uso del amplio instrumental relativo a la persistencia y cointegración entre series temporales económicas[43]. Es necesario partir desde un principio de la definición de conceptos como equilibrio y convergencia (Fingleton (1997)), que pueden ser diferentes según que el contexto en el que nos movamos sea estacionario o no, y el concepto de β-convergencia dado al principio de esta sección parece tener en mente un mundo estacionario, mientras que la realidad puede ser muy diferente. De esta forma la definición de ***f*-convergencia** dada anteriormente **puede acomodarse con facilidad a la presencia de raíces unidad**, lo que sin embargo no soluciona la cuestión de que inferencias teóricas podemos derivar de estas posibles regularidades empíricas.

[42] Dejando al margen las posibles complicaciones derivadas de que los procesos de crecimiento pueden mostrar importantes **nolinealidades** y **multiplicidad del equilibrio** (Azariadis y Drazen (1990)), que algunos autores (Durlauf y Johnson (1995), Hansen (2000)) han señalado como potencialmente importantes.
[43] La tecnología para aplicar estas técnicas en contextos de datos de panel está todavía en su infancia (Phillips y Moon (1999)).

β*-convergencia* **versus** **σ*-convergencia.***

Ya hemos indicado como el concepto de **β-convergencia** trata de examinar si las **economías inicialmente pobres**, con bajos niveles de renta *per capita* en términos relativos, **han tendido a crecer más que las economías inicialmente ricas**, con altos niveles de renta *per capita.* Debido a ello podríamos pensar que si en una muestra concreta encontramos β-convergencia entonces, debe haberse producido una reducción en la dispersión *cross-section* de la renta *per capita,* en otras palabras debe haberse producido una reducción en la desigualdad en la distribución de la renta; de forma que β-convergencia implica σ-convergencia. Sin embargo es bien conocido que esta relación no tiene porque cumplirse y la existencia de **β-convergencia es compatible con diferentes situaciones en términos de σ-convergencia**.

La forma más simple e intuitiva de darse cuenta de ello es la siguiente. Supongamos que

- **(*i*) para cada economía, *i*, x_{it} es un proceso estocástico estacionario** indexado por t y con segundos momentos finitos, dicho proceso estocástico es idéntico para cada i;
- **(*ii*) para cada *t*, x_{it} es una colección de variables aleatorias independientes e idénticamente distribuidas**, un proceso estocástico indexado por i que podríamos definir como un ruido blanco en el corte transversal, esto se verifica para cada t.

Bajo estos dos supuestos, llamando $\sigma_t^2 = Var(\log x_{it})$ a la dispersión *cross-section*[44] y tomando varianzas a ambos lados de (28) obtenemos

$$\sigma_t^2 = \rho^2 \sigma_{t-1}^2 + \sigma_u^2 \qquad (37)$$

[44] Esta es una razón para la popularidad de $Var(\log x_{it})$ como medida de dispersión *cross-section*, ciertamente en el contexto de (28) esta parece ser una medida natural si bien ya hemos observado en Goerlich (2000a) como este estadístico no es el que mejores propiedades tiene. En esta sección utilizaremos $Var(\log x_{it})$ como medida de dispersión simplemente porque nos permite obtener resultados exactos de forma sencilla y por tanto esperamos que se verifiquen aproximadamente para otras medidas de dispersión, sin embargo, es necesario tener presente que $Var(\log x_{it})$ no verifica el principio de las transferencias de Pigou (1912)-Dalton (1920), lo que puede llevar a situaciones curiosas (Foster y Ok (1999)). En la práctica es importante que medida de dispersión utilicemos.

siendo $\sigma_u^2 = Var(u_{it})$, que se supone invariante en el tiempo, puesto que (*i*) y (*ii*) implican $E(u_{it}) = 0$ y $Cov(\log x_{it-1}, u_{jt}) = 0$ $\forall i,j,t$.

Sustituyendo recursivamente *j* periodos hacia atrás en (37) obtenemos

$$\sigma_t^2 = \rho^{2j}\sigma_{t-j}^2 + \sigma_u^2 \Sigma_{k=0}^{j-1}\rho^{2k} \tag{38}$$

Por lo tanto la existencia de β-convergencia, $0 < \beta < 1 \Rightarrow 0 < \rho < 1$, implica que, con el paso del tiempo, conforme $j \to \infty$

$$\sigma^2 = \lim_{j\to\infty} \sigma_t^2 = \frac{\sigma_u^2}{1-\rho^2} \tag{39}$$

en consecuencia la **dispersión de la distribución estacionaria** de $\log x$ es $\sigma^2 = \frac{\sigma_u^2}{1-\rho^2}$, dispersión hacia la que tiende σ_t^2 de forma monótona.

En consecuencia, a pesar de la existencia de β-convergencia, si obtenemos además σ-convergencia ello depende de que la dispersión inicial, digamos σ_0^2, sea mayor que $\sigma^2 = \frac{\sigma_u^2}{1-\rho^2}$. Esto representa en realidad una restricción sobre el periodo inicial, si $\sigma_0^2 > \frac{\sigma_u^2}{1-\rho^2}$, entonces σ_t^2 debe disminuir de forma continuada hasta su valor límite y observaremos **σ-convergencia**; por el contrario si $\sigma_0^2 < \frac{\sigma_u^2}{1-\rho^2}$, entonces σ_t^2 debe aumentar de forma monótona hasta su valor de estado estacionario y observaremos **σ-divergencia**; finalmente, si dicho estado estacionario ya ha sido alcanzado, de forma que $\sigma_0^2 = \frac{\sigma_u^2}{1-\rho^2}$, entonces σ_t^2 ya ha convergido, la distribución *cross-section* ha alcanzado su estado estacionario, y observaremos **s-constancia**[45]. Por tanto, **b-convergencia no implica s-convergencia** (Barro y Sala-i-Martín (1992), p.-227-228), **es decir b-convergencia es una condición necesaria pero no suficiente para obtener s-convergencia**[46].

El conjunto de gráficos 2 (a)-(c) permite ilustrar las diversas situaciones una vez eliminada la tendencia creciente en el estado estacionario debido a la existencia de progreso técnico, todas ellas partiendo de la misma situación de β-convergencia, $0 < \beta < 1 \Rightarrow 0 < \rho < 1$. En el **gráfico 2 (a)** las economías parten de una situación en la que están relativamente concentradas entorno a un punto y conforme pasa el tiempo la dispersión se aproxima a su valor de estado estacionario, puesto que la dispersión inicial es menor que la de la distribución estacionaria observaremos σ-divergencia. Por el contrario el **gráfico 2 (b)** ilustra la situación contraria, la dispersión inicial es mayor que la correspondiente al estado estacionario y conforme pasa el tiempo la dispersión disminuye, de esta forma observaremos σ-convergencia. Esta es la situación que parecen tener en mente muchos estudiosos aplicados de la literatura sobre convergencia, o al menos la situación que les gustaría observar en la práctica, aquella en la que **b-convergencia y s-convergencia coinciden**, ya que en este caso parece razonable hablar de las economías pobres dando alcance (*catching-up*) a las economías ricas, al menos en un sentido promedio, sin embargo nuestros razonamientos ilustran que esto es una mera posibilidad, y no la única que podemos encontrar en la práctica.

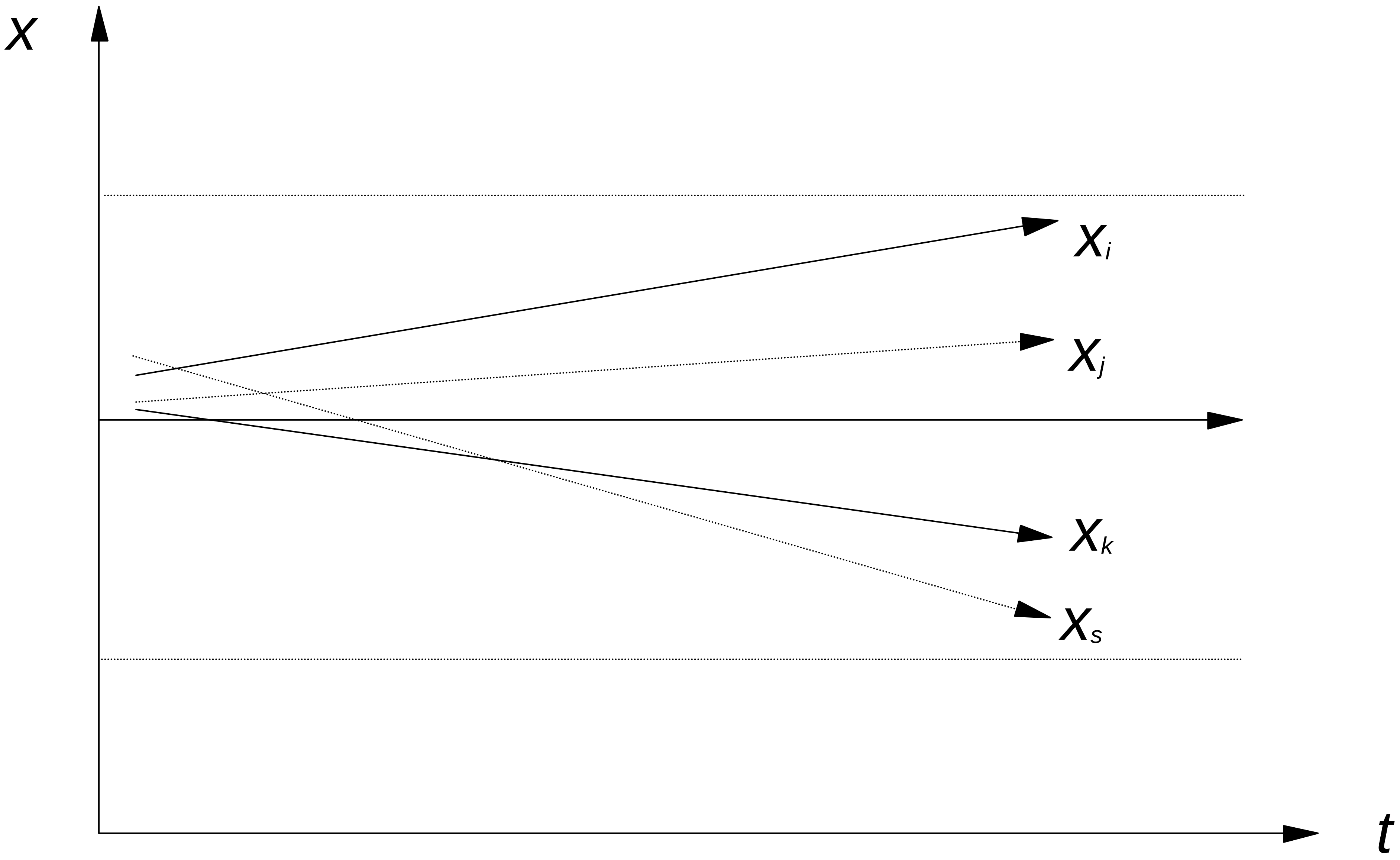
Gráficos 2 a,b,c,d

Finalmente el **gráfico 2 (c)** ilustra una situación en la que el estado estacionario ya ha sido alcanzado, la dispersión permanece constante pero existe una notable movilidad intradistribucional de forma que las economías intercambian sus posiciones relativas con frecuencia, todo esto sucede con σ-constancia, pero las economías ricas crecen menos que las pobres de forma que observaremos β-convergencia, aunque en este caso con

[45] Es necesario recordar que estamos razonando en términos poblacionales, en términos muestrales debemos observar aproximádamente estas características de la población.

[46] Obsérvese que β-divergencia implica necesariamente un incremento de la dispersión, lo que justifica la necesidad de β-convergencia para obtener σ-convergencia.

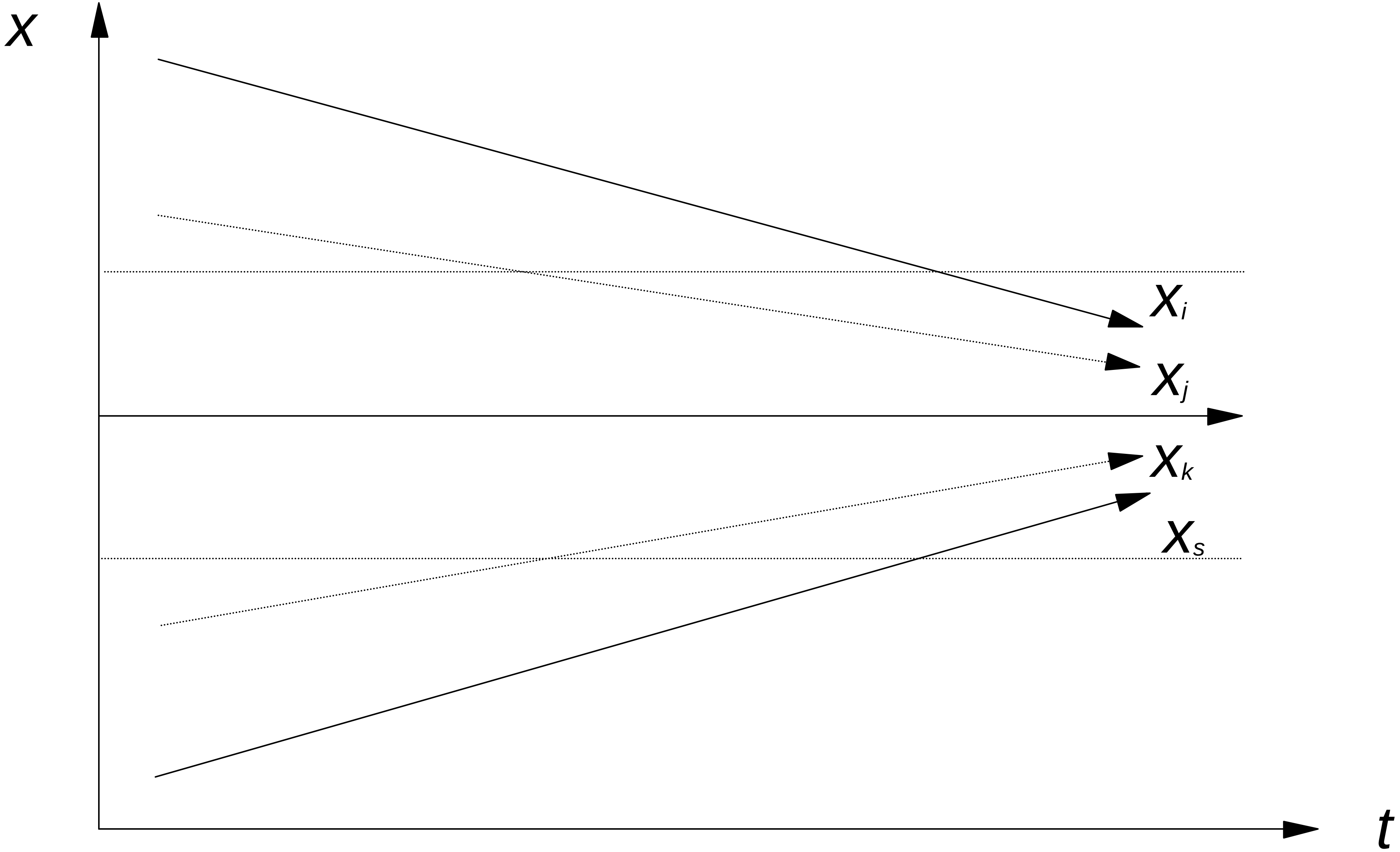
Gráfico 2 (a). *Sigma*-divergencia



Sigma-divergencia con dispersión estacionaria en el largo plazo

Las economías comienzan relativamente juntas respecto a la situación de estado estacionario y conforme transcurre el tiempo su dispersión aumenta hasta converger en distribución a un estado estacionario bien definido.

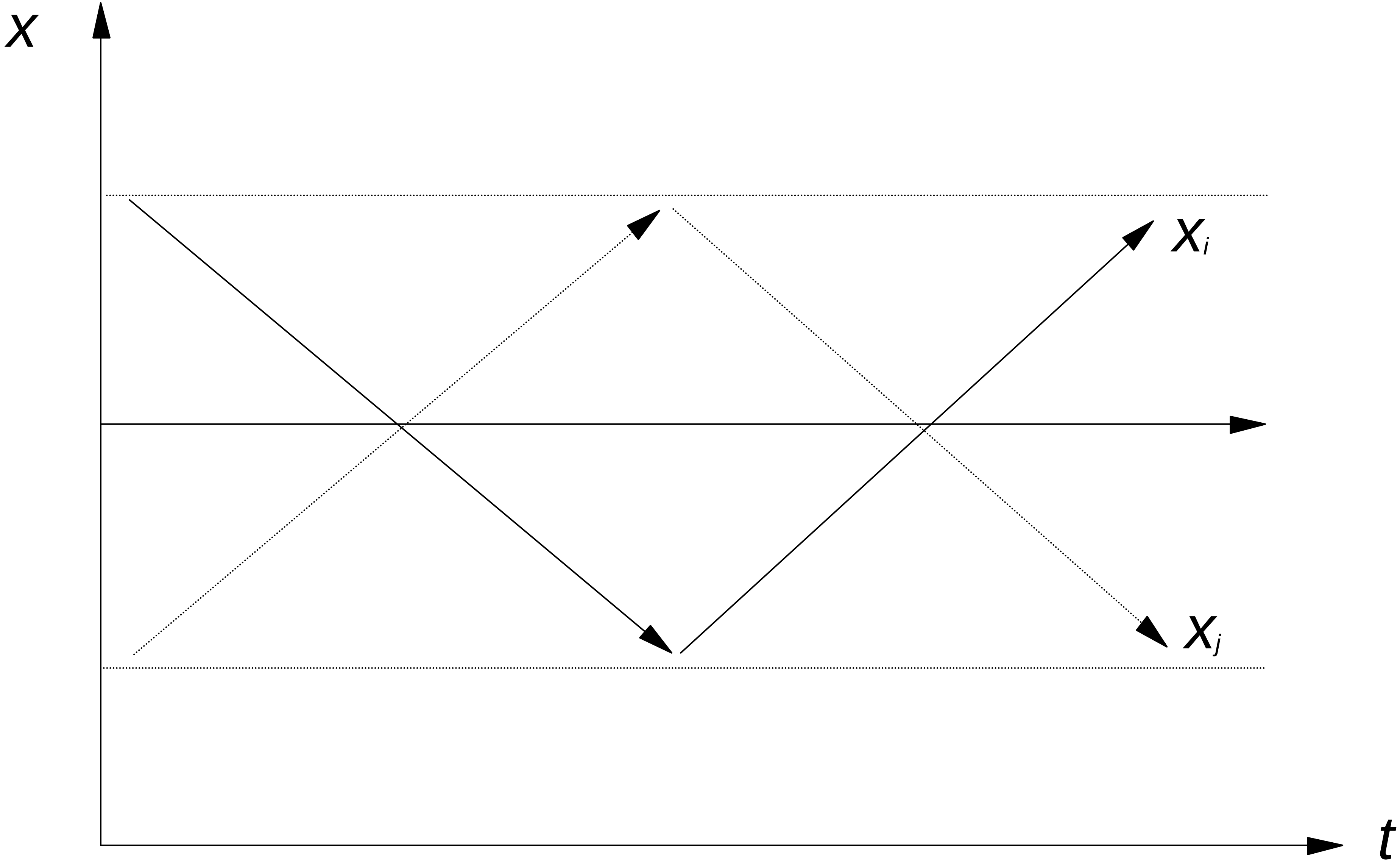
Gráfico 2 (b). *Sigma*-convergencia



Sigma-convergencia con dispersión estacionaria en el largo plazo

Sigma-convergencia y Beta-convergencia coinciden. Las economías, inicialmente dispersas con respecto al estado estacionario, convergen monotónicamente hacia dicho estado; durante la transición la dispersión disminuye.

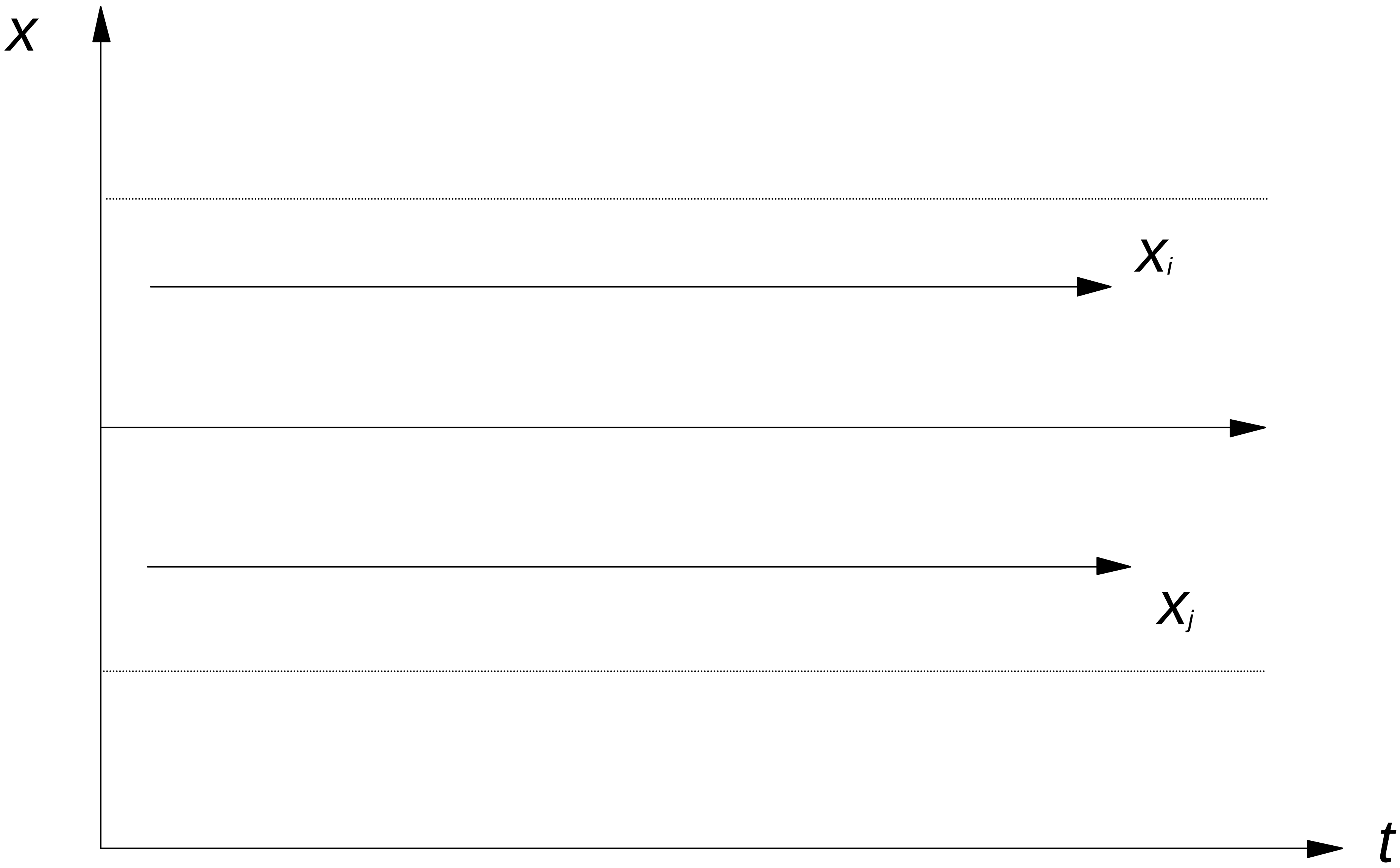
Gráfico 2 (c). *Sigma*-constante y gran movilidad



Sigma-constante con dispersión estacionaria en el largo plazo y gran movilidad

Las economías han alcanzado ya el estado estacionario y su dispersión es constante, pero existe una elevada movilidad intra-distribucional y las economías individuales intercambian sus posiciones relativas a lo largo del tiempo.

Gráfico 2 (d). *Sigma*-constante y gran persistencia



Sigma-constante con dispersión estacionaria en el largo plazo y gran persistencia

Las economías han alcanzado ya el estado estacionario y su dispersión es constante, pero existe una persistencia extrema, de forma que las economías individuales mantienen sus posiciones relativas a lo largo del tiempo.
Las economías simplemente se mueven en paralelo.

$1 < \beta < 2 \;\Rightarrow\; -1 < \rho < 0$. Por último el **gráfico 2 (d)** ilustra una situación similar a la del gráfico 2 (c), en el sentido de que el estado estacionario ya ha sido alcanzado y por tanto la dispersión permanece constante, pero al contrario que en 2 (c) ahora existe completa persistencia, las economías ricas permancecen ricas y las economías pobres permanecen pobres, en este caso no observaremos β-convergencia. Aunque la información en términos de σ-convergencia en los gráficos 2 (c) y (d) es idéntica la situación es en realidad muy diferente. Son estos razonamientos los que permiten argumentar a Quah (1993a,b, 1996e) que el concepto de β-convergencia, o en general el estudio de las ecuaciones de convergencia *cross-section* o mediante técnicas de datos de panel (Durlauf y Quah (1998)), son completamente inútiles en términos de estudiar la dinámica de las distribuciones en el tiempo. Las regresiones *cross-section* representan el comportamiento medio de un conjunto de economías pero no el comportamiento de la distribución y los argumentos anteriores ponen de manifiesto que **lo que sucede con la media condicional**, que es lo que representa una regresión *cross-section*, **no es muy útil en términos de saber que es lo que pasa con la totalidad de la distribución**.

Obsérvese que el papel de σ_u^2 en (39) es importante en el razonamiento anterior, desde el punto de vista de la teoría del crecimiento u_{it} se interpreta como una perturbación transitoria, si esta no existiera entonces $\sigma_u^2 = 0$ y $\sigma^2 = \lim_{j\to\infty}\sigma_t^2 \to 0$, en este caso deberíamos observar el colapso de los niveles de renta *per capita* a un solo punto y β-convergencia si implicaría σ-convergencia, pero ya hemos indicado como este caso no es útil en la práctica aunque volveremos brevemente sobre él mas adelante. Por otra parte es necesario interpretar estas perturbaciones sobre la distribución *cross-section* como un continuo en el tiempo de forma que las inferencias que podemos extraer de los datos son mucho más complejas de lo podemos deducir a partir del conjunto de gráficos 2 (a)-(d).

La **relación entre** los modernos conceptos de **β-convergencia y σ-convergencia** se remonta al origen mismo de la regresión (Galton (1877)), cuya historia será comentada brevemente en el epígrafe siguiente, y ha causado numerosas confusiones desde entonces (Secrist (1933), Baumol, Blackman y Wolff (1989), Williamson (1991)), a pesar de que la relación subyacente entre ambos conceptos es bien entendida (Hotelling (1933), Hart y Prais (1956), Prais (1958), Friedman (1992), Quah (1993a), Hart (1995)). En el contexto de nuestros supuestos (*i*) y (*ii*) dicha relación puede ser formalmente establecida de la siguiente forma (Quah (1993a, p.-432), Durlauf y Quah (1998, p.-40)). Llamemos $y_{it} = \log x_{it}$ por simplicidad y supongamos, de acuerdo con (*i*) y (*ii*), que $(y_{i,t-1}, y_{it})'$ es normal bivariante estacionaria[47] para todo *i*; en consecuencia muestras extraidas de una población con estas características deben mostrar dispersión constante, es decir σ-constancia, ¿que debemos esperar en términos de β-convergencia?. Observamos que (28) puede ser escrita como

$$E(y_{it}|y_{i,t-1}) = \mu + \rho(y_{i,t-1} - \mu) \tag{40}$$

siendo $E(y_{it}) = E(y_{i,t-1}) = \mu \ \ \forall t$ por estacionariedad[48] y $\rho = \dfrac{Cov(y_{it}, y_{i,t-1})}{Var(y_{i,t-1})}$.

La **desigualdad de Cauchy-Schwarz** (Spanos (1999), p.-275) implica que

$$\vartheta^2 = Cor^2(y_{it}, y_{i,t-1}) = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var(y_{i,t-1})} \leq 1 \tag{41}$$

mientras que estacionariedad implica $Var(y_{it}) = Var(y_{i,t-1}) = \sigma^2 \ \ \forall t$, en consecuencia ambos resultados nos dan la siguiente relación

$$\vartheta^2 = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var^2(y_{it})} = \rho^2 \leq 1 \quad \Rightarrow \quad |\rho| < 1 \ \text{ si } \ |\vartheta| < 1 \tag{42}$$

Por lo tanto **un signo negativo en el coeficiente sobre la condición inicial en la regresión *cross-section* de convergencia no indica una reducción en la dispersión o desigualdad**, ya que como hemos observado en este ejemplo la dispersión permanece constante

---

[47] El supuesto de normalidad no es necesario, ya que el argumento puede ser racionalizado en términos de proyecciones lineales (Quah (1993a)), sin embargo facilita algunos cálculos.
[48] α en (28) es por tanto $\mu(1-\rho)$ en (40).

pero con tal que la correlación entre la situación inicial y final no sea idénticamente igual a 1 (gráfico 2 (d)) obtendremos β-convergencia[49].

Este es el argumento que permite a Quah (1993a) afirmar que las regresiones *cross-section* son completamente inútiles en términos de obtener conclusiones sobre la evolución dinámica de la distribución en el corte transversal. Una opinión contraria, que tiende a resaltar la importancia del concepto de β-convergencia, puede encontrarse en Sala-i-Martín (1994, 1996) pero obsérvese que en sus ejemplos siempre se utilizan variables en términos de *rankings*, los equipos de *football* en una liga, y en este caso la distribución de estas variables es claramente estacionaria con media y varianza constante y por tanto $|\vartheta| \leq 1$ con igualdad positiva si y sólo si todos los *rankings* se mantienen, es decir todo permanece igual. Puesto que en este caso estacionario ya hemos observado que $\vartheta = \rho$, obtendremos siempre β-convergencia a no ser que todas las observaciones mantengan su posición relativa, es decir en este caso concreto β-convergencia esta inexorablemente asociado a la existencia de movilidad intradistribucional, claramente **la movilidad dentro de la distribución está relacionada con el concepto de β-convergencia, pero ambos conceptos no son equivalentes**, de hecho el coeficiente de correlación entre la situación inicial y final, ϑ, puede ser considerado como una medida de movilidad intradistribucional, especialmente si la variable está medida en términos de *rankings* (Goerlich (2001a)), pero sólo en el caso en que $(y_{i,t-1}, y_{it})'$ sea estacionaria se cumple que $\vartheta = \rho = 1 - \beta$. Por tanto en los ejemplos utilizados por Sala-i-Martín (1994, 1996) sólo es posible obtener β-convergencia si hay movilidad intradistribucional, no siendo posible una β-convergencia monótona hacia el estado estacionario como la ilustrada en el gráfico 2 (b), que es precisamente la que desde un punto de vista teórico predice el modelo de crecimiento neoclásico; de hecho la condición $\beta < 1$, mencionada anteriormente, excluye alteraciones en las posiciones relativas entre las observaciones (*leapfrogging*), es decir excluye movilidad[50].

El resultado que acabamos de mostrar es un resultado muy potente, pero aún podemos ser más explícitos acerca de la relación entre β-convergencia y σ-convergencia. Por ejemplo

---

[49] Tampoco obtendríamos β-convergencia si la correlación entre la situación inicial y final fuera -1.

[50] Otros autores han tendido a identificar el concepto de β-convergencia con medidas estadísticas de movilidad intradistribuciónal (Boyle y McCarthy (1997)), lo que es claramente incorrecto.

nada en el argumento anterior cambia si relajamos el supuesto de idéntica distribución en el corte transversal, el caso más sencillo es simplemente si permitimos que las diferentes economías tengan estados estacionarios distintos, $E(y_{it}) = \mu_i \;\; \forall t$, de forma que (40) se convierte en

$$E(y_{it}|y_{i,t-1}) = \mu_i + \rho(y_{i,t-1} - \mu_i) \tag{43}$$

pero nada cambia en los resultados puesto que de nuevo $\vartheta^2 = \dfrac{Cov^2(y_{it}, y_{i,t-1})}{Var^2(y_{it})} = \rho^2 \leq 1$.

Podemos igualmente permitir un mayor grado de heterogeneidad en el corte transversal[51], por ejemplo $Var(y_{it}) = \sigma_i^2 \;\; \forall t$, o incluso permitir cierto grado de dependencia, débil o fuerte, entre las diferentes economías (Quah (1993a), p.-434), aunque el álgebra es más complicada los resultados esenciales se mantienen. El supuesto (*ii*) es meramente simplificador pero en modo alguno esencial para nuestra **conclusión**, **b-convergencia no implica s-convergencia, es decir b-convergencia es una condición necesaria pero no suficiente para obtener s-convergencia**; además dicho resultado no descansa sobre el supuesto de independencia e idéntica distribución en el corte transversal.

Este resultado ha sido demostrado en el contexto de la estacionariedad del proceso $(y_{i,t-1}, y_{it})'$ pero ¿que sucede si relajamos este requisito y permitimos que $Var(y_{it})$ varíe en el tiempo?, **¿podemos relajar el supuesto (*i*) sin que ello altere nuestra conclusión fundamental?**, la respuesta es **si**, y además encontramos entonces una **relación** interesante **entre b-convergencia y s-convergencia** (Hart y Prais (1956), Prais (1958), Hart (1995)). A partir de la definición del coeficiente de correlación, $\vartheta^2 = \dfrac{Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var(y_{i,t-1})}$, y del coeficiente de β-convergencia en (28), $\rho = \dfrac{Cov(y_{it}, y_{i,t-1})}{Var(y_{i,t-1})}$, es fácil obtener la siguiente relación

---

[51] Con heterogeneidad en el corte transversal cualquier distribución *cross-section* invariante en el tiempo (estacionaria) es una mezcla probabilística de las diferentes distribuciones temporales individuales.

$$\vartheta^2 = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var(y_{i,t-1})} = \frac{Var(y_{i,t-1}).Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var^2(y_{i,t-1})} = \frac{Var(y_{i,t-1})}{Var(y_{it})}.\rho^2 \tag{44}$$

por tanto,

$$\frac{Var(y_{it})}{Var(y_{i,t-1})} = \frac{\rho^2}{\vartheta^2} \tag{45}$$

lo que nos permite obtener la siguiente **tipología** de casos

(1)	$\rho^2 > 1$	$\Rightarrow$	$Var(y_{it}) > Var(y_{i,t-1})$	**s-divergencia**
(2)	$\vartheta^2 < \rho^2 < 1$	$\Rightarrow$	$Var(y_{it}) > Var(y_{i,t-1})$	**s-divergencia**
(3)	$\rho^2 < \vartheta^2 < 1$	$\Rightarrow$	$Var(y_{it}) < Var(y_{i,t-1})$	**s-convergencia**

En consecuencia **la condición** $\rho^2 < 1$, o alternativamente $0 < \beta < 2$, **por sí sola no nos permite alcanzar ninguna conclusión acerca de la evolución en la dispersión de la distribución *cross-section*, aunque** $\rho^2 > 1$ **si indica de forma inequívoca s-divergencia**.

Obsérvese que ϑ es la **correlación entre la situación inicial y final**, $(y_{i,t-1}, y_{it})$, un estadístico que, como ya hemos mencionado, puede ser utilizado como medida de movilidad intradistribucional, un **valor bajo** de ϑ implica **poca correlación entre la situación inicial y la final**, y por tanto tiende a indicar una **elevada movilidad** dentro de la distribución de $\log x_{it}$ a lo largo del tiempo, por el contrario un **valor alto** de ϑ implica una **elevada correlación entre la situación inicial y la final**, y por tanto tiende a indicar una **baja movilidad** dentro de la distribución de $\log x_{it}$ a lo largo del tiempo. **Si la movilidad es suficientemente baja**, de forma que $\rho^2 < \vartheta^2 < 1$, **entonces obtenemos s-convergencia**, $Var(y_{it}) < Var(y_{i,t-1})$.

En **conclusión**, **cuando** $\rho^2 < 1$ **si la movilidad entre economías**, u observaciones en general, **es baja en relación a la magnitud de la relación entre crecimiento y condición inicial**, las economías inicialmente pobres creciendo más que las inicialmente ricas, **entonces la**

dispersión se reduce y encontraremos una situación de s-convergencia, pero **por el contrario si la movilidad dentro de la distribución es alta en relación a la magnitud de la relación entre crecimiento y condición inicial entonces la dispersión aumenta y obtendremos s-divergencia**.

Esto explica **porque b-convergencia es popular en la práctica**, al menos en el contexto de la teoría del crecimiento, en **primer lugar** porque **es intuitiva**, y en **segundo lugar** porque, a pesar de que esta intuición puede a veces ser engañosa en términos de lo que nos indica acerca de la dispersión en la distribución *cross-section*, **coincide**, en muchas ocasiones, **con s-convergencia**, ya que no es de esperar que la movilidad sea especialmente alta, ni en términos de economías regionales o de países ni tampoco en términos de la distribución personal de la renta (Zimmerman (1992), Cantó (2000)).

La **descomposición de los cambios temporales en la dispersión de la renta *per capita* en los dos efectos** que nos proporciona el anterior resultado, $\frac{Var(y_{it})}{Var(y_{i,t-1})} = \frac{\rho^2}{\vartheta^2}$, lo que podríamos denominar el **efecto "regresion"**, capturado por ρ, y el denominado **efecto movilidad**, capturado por ϑ, no parece que haya sido explotada por la literatura del crecimiento económico, aunque si por la literatura dedicada al análisis de la concentración industrial (Hart y Prais (1956), Prais (1958), Davis, Haltiwanger y Schuh (1993)).

Finalmente un comentario marginal, obsérvese que el anterior argumento acerca de resultados contrarios entre β-convergencia y σ-convergencia requiere cierto grado de aleatoriedad entre las situaciones inicial y final, es decir exige que $\vartheta^2 < 1$, cuando dicha relación es de carácter determinista entonces $\vartheta^2 = 1$ y no existe conflicto entre β-convergencia y σ-convergencia, en el sentido de que $\rho^2 > 1$, $\beta < 0$ ó $\beta > 2$, indica σ-divergencia y al mismo tiempo $\rho^2 < 1$, $0 < \beta < 2$, indica inequívocamente σ-convergencia (Prais (1958), p.-269).

Galton (1877) y la historia de la regresión.

falacia

Engaño, fraude o mentira con que se intenta dañar a otro.
paradoja
Idea extraña u opuesta a la común opinión y al sentir de los hombres.
Diccionario de la Real Academia Española.

El hecho de que β-convergencia y σ-convergencia no se impliquen mutuamente es un hecho bien conocido en estadística y que se remonta al **origen** mismo **de la regressión**, concepto que fue propuesto inicialmente de forma tímida por Galton (1877), formalizado por Galton (1885, 1886a,b) con la ayuda de Dickson (1886), desarrollado posteriormente por Pearson (1894, 1895, 1896) y finalmente relacionado con la tradición actual de mínimos cuadrados ordinarios por Yule (1897)[52]. Dicho hecho es etiquetado muchas veces como **"la falacia de Galton"** (Friedman (1992), Quah (1993a)), pero como acabamos de ver no existe nada de engañoso en la relación entre ambos conceptos, si acaso **paradoja** sería un término más acertado.

Francis Galton, nacido el 16 de febrero de 1822 y fallecido el 17 de enero de 1911, fue un estadístico notable (Galton (1908), Fisher (1956), MacKenzie (1981), Porter (1986), Stigler (1986)) al que su interés por el estudio de la dependencia entre variables le llevó de forma natural a la consideración de distribuciones conjuntas y condicionadas, y partir de ellas a dos de los conceptos que más influencia han tenido en la estadística moderna, la **regresión** (Galton (1886a)) y la **correlación** (Galton (1888)).

La noción de regresión fue propuesta por Galton (1877) en el contexto del estudio de las características hereditarias de dos generaciones de guisantes de olor[53] y dicha noción fue inicialmente denominada **reversión**. Sin embargo el trabajo que estableció la regresión

[52] Existe poco de los origenes en la concepción actual de la regresión (Maddala (1977), p.-97-101), que curiosamente ha mantenido su terminología, a pesar de que no poseer hoy en día ninguna relación con la noción de regresar, acción de retroceder o volver hacia atrás (Anscombe (1967)).
[53] El guisante de olor es una planta aromática, variedad de almorta, que se cultiva en los jardines, tiene flores amariposadas, tricolores y de excelente perfume y es además muy trepadora. El interés de Galton por la biología estuvo influido probablemente por el ambiente familiar, era nieto de Erasmus Darwin y primo de Charles Darwin.

propiamente dicha fue Galton (1886a) en el contexto del estudio de la relación entre la altura media de los padres y la altura de sus hijos.

Galton (1886a) encontró que existía una tendencia a que los padres de estatura elevada tuvieran hijos altos y que los padres de estatura baja tuvieran hijos bajos. De esta forma, examinando la distribución empírica de las alturas de una generación y la siguiente, Galton (1886a) observó que la distribución de alturas, estudiada mediante histogramas, parecía permanecer estable, en concreto la misma curva de frecuencias normal parecía describir la distribución de alturas de padres e hijos, de forma que dicha distribución parecía replicarse a sí misma generación tras generación, al igual que en el caso de las características de los guisantes de olor (Galton (1877)). Sin embargo al mismo tiempo Galton (1886a,b) encontró, a partir de sus observaciones, que la estatura promedio de los hijos de padres con una determinada altura tendía a "regresar" o "revertir" hacia la estatura media del total de la población. Galton (1886a) llegó a esta conclusión de forma empírica, trazando la recta de regresión que proporcionaba la altura media de los hijos condicionada en una altura dada de los padres y observando que esta recta tenía una pendiente menor que la unidad[54], Galton (1886a) llamó a este fenómeno "*regression towards mediocrity*". De esta forma Galton (1886b) no sólo derivó la distribución normal bivariante con la ayuda de Dickson (1886), sino también las rectas que definen a partir de esta distribución las esperanzas condicionadas de una variable respecto a la otra.

La pregunta que se hizo Galton (1886a) fue la misma que se había hecho algunos años antes estudiando las características hereditarias de los guisantes de olor:

> "*How is it that although each individual does not as a rule leave his like behind him, yet successive generations resemble each other with great exactitude in all their general features?...*"
>
> Galton (1877, p.-492)

---

[54] El diagrama basado en la tabulación de la Tabla I de Galton (1986a) muestra gráficamente la derivación de las rectas de regresión a partir de la elipses que definen la distribución empírica bivariante de alturas de padres e hijos.

o dicho en nuestra terminología, como es que obtenemos β-convergencia pero no σ-convergencia, sino por el contrario la dispersión se mantiene constante generación tras generación. La respuesta ha sido dada en el epígrafe anterior, si la distribución es estacionaria, tal y como Galton (1885, 1886a,b) supuso de forma implícita, entonces las ecuaciones (40) y (41) implican que $\vartheta^2 = \rho^2 \leq 1$, es decir $|\rho| < 1$ si $|\vartheta| < 1$, por tanto con tal que las correlaciones entre alturas de padres e hijos no sea perfecta encontraremos el fenómeno de regresión a la media, en el sentido de que padres con alturas muy elevadas tendrán hijos que, por término medio, no tendrán una altura tan elevada como la de sus padres y al mismo tiempo padres con alturas muy bajas tendrán hijos que, por término medio, no serán tan bajos como sus padres. Galton (1886a) hizo de su "*regression towards mediocrity*" una ley de herencia genética (Galton (1869, 1877, 1889)) subyacente a la aparente estabilidad de las características de la población en generaciones sucesivas e incorrectamente extrajo conclusiones de causalidad a partir de su recta de regresión[55].

¿Condicionar en el pasado o en el futuro?

Hemos observado al principio de esta sección como el concepto de **b-convergencia trata de examinar si las economías inicialmente pobres**, con bajos niveles de renta *per capita* al principio del periodo, **han tendido a crecer más que las economías inicialmente ricas**, con altos niveles de renta *per capita* al principio del periodo. Puesto que β-convergencia implica comparar dos momentos del tiempo podemos **invertir la perspectiva temporal** y preguntarnos igualmente si **las economías finalmente ricas**, con altos niveles de renta *per capita* al final del periodo, **han tendido a crecer más que las economías finalmente pobres**, con bajos niveles de renta *per capita* al final del periodo. En este sentido podemos definir el concepto de γ-convergencia entre un conjunto de unidades económicas, países, regiones o individuos, si existe una **relación positiva entre la tasa de crecimiento** de la renta *per capita* (o cualquier otra variable) de dichas unidades económicas y su **valor final**. De nuevo este es un fenómeno de **'regresión o reversión a la media'**. Al igual que β-convergencia, **g-**

[55] Obsérvese que sus argumentos estadísticos son simétricos respecto a la recta de regresión de padres a hijos o de hijos a padres, algo de lo que hablaremos en el epígrafe siguiente; lo que muestra la imposibilidad de extraer conclusiones causales de simples ejercicios de condicionamiento.

convergencia es un concepto dinámico que **relaciona la situación final con el crecimiento previo de una variable**.

En un mundo lineal y sujeto a incertidumbre podemos formalizar la idea de γ-convergencia mediante la ecuación

$$g_{x_i} = -\alpha' + \gamma x_{it} - u'_{i,t-1} \qquad (46)$$

donde g_{x_i} representa la tasa de crecimiento de la renta per capita, x_{it} la condición final y $u'_{i,t-1}$ un término de perturbación que captura *shocks* transitorios (estacionarios) sobre la tasa de crecimiento del individuo o región i y que como primera aproximación podemos suponer independiente e idénticamente distribuido, tanto en el corte transversal como en la dimensión temporal. La existencia de **g-convergencia implica g > 0 en (46)**, puesto que en este caso la tasa de crecimiento de x, g_x, está positivamente relacionada con la condición final, x_t.

Podemos hacer ahora, a partir de (46), un análisis simétrico para la γ-convergencia al ya realizado para el concepto de β-convergencia en los epígrafes anteriores. En particular adoptaremos como **especificación operativa** para cuantificar el concepto de γ-convergencia una ecuación logarítmico-lineal

$$\log x_{it} - \log x_{i,t-1} = -\alpha' + \gamma \log x_{it} - u'_{i,t-1} \qquad (47)$$

que puede ser convenientemente escrita como

$$\begin{aligned} \log x_{i,t-1} &= \alpha' + (1-\gamma)\log x_{it} + u'_{i,t-1} \\ &= \alpha' + \rho' \log x_{it} + u'_{i,t-1} \qquad \rho' = 1-\gamma \end{aligned} \qquad (48)$$

un proceso futurista AR(1) en logaritmos en el que el presente está escrito en función del futuro; por lo tanto **g-convergencia, g > 0, implica r¢ < 1** en (48), más concretamente si nos restringimos a lo que podíamos llamar, por similaridad, la situación más habitual,

$$0 < \gamma < 1 \quad \Leftrightarrow \quad 0 < \rho' < 1$$

lo que indica que γ-convergencia, en el sentido que lo hemos definido, implica, en términos de series temporales, un proceso *forward* estacionario con autocorrelación positiva para los logaritmos de x y en el que el presente está en función del futuro.

Es bien conocido que los procesos estocásticos estacionarios pueden ser revertidos (Kim (1997)) de forma que la flecha del tiempo puede verse como del pasado al futuro o alternativamente del futuro al pasado y por tanto (48) contiene la misma información que (28), o alternativamente (46) la misma que (14). En consecuencia **los conceptos de b-convergencia y g-convergencia no son en realidad conceptos diferentes sino dos formas alternativas de examinar la misma realidad y proporcionan informaciones complementarias**.

En el estudio sobre las tasas de crecimiento entre economías **¿debemos condicionar en el pasado o en el futuro?**[56], es decir ¿la ecuación de convergencia debe realizarse sobre la condición inicial, (14), tal y cómo normalmente se hace, o sobre la condición final, (46)?. En el contexto de la teoría del crecimiento esta cuestión aparece marginalmente sólo en Quah (1993a) y Hart (1995), quienes ofrecen respuestas diferentes a la pregunta anterior, sin embargo fuera de nuestro contexto la cuestión aparece con frecuencia en la literatura sobre concentración en economía industrial (Hart y Prais (1956), Davis, Haltiwanger y Schuh (1993)) y ya fue objeto de mención en los origenes mismos de la regresión (Galton (1886a)).

La **respuesta** de si debemos condicionar en el pasado o en el presente es simple. Desde un punto de vista estadístico es **absolutamente indiferente**, es decir dado el proceso $(y_{i,t-1}, y_{it})'$, examinado al analizar la relación entre β-convergencia y σ-convergencia, resulta indiferente estimar la ecuación (43) o alternativamente

$$E(y_{i,t-1} | y_{it}) = \mu_i + \rho'(y_{it} - \mu_i) \tag{49}$$

[56] La misma pregunta podría realizarse respecto a cualquier punto intermedio entre $t-1$ y t (Quah (1993a)).

donde $\rho' = \frac{Cov(y_{it}, y_{i,t-1})}{Var(y_{it})}$. Es fácil observar que $\rho\rho' = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var(y_{i,t-1})} = \vartheta^2 \leq 1$, lo que clarifica la **relación entre los conceptos de β-convergencia y γ-convergencia**.

En concreto obsérvese que puesto que $\rho\rho' \leq 1$ no podemos obtener simultáneamente que $\rho > 1$ y $\rho' > 1$, es decir, β-divergencia junto con γ-divergencia; por el contrario si podemos observar $\rho < 1$ y $\rho' < 1$, es decir, β-convergencia junto con γ-convergencia; esta situación la observaremos, por ejemplo, en el **caso estacionario**,

$$Var(y_{it}) = Var(y_{i,t-1}) \quad \forall t \qquad \Rightarrow \qquad \rho = \rho'$$

en el que $\rho^2 = \rho'^2 = \vartheta^2 \leq 1$, y por tanto $0 < \rho < 1$ implica $0 < \rho' < 1$. Adicionalmente podemos observar $\rho < 1$ y $\rho' > 1$, es decir, β-convergencia junto con γ-divergencia; o alternativamente $\rho > 1$ y $\rho' < 1$, es decir, β-divergencia junto con γ-convergencia.

Ciertamente la elección entre condicionar en el pasado o en el futuro importa si deseamos interpretar los parámetros de nuestra ecuación, una esperanza condicional, en términos teóricos de algún modelo (económico) subyacente, en este caso deberemos suponer que la variable condicionante está dada en el momento en que la condicionada se determina de forma que, puesto que en el mundo real la flecha del tiempo es unidireccional, condicionar en el pasado será el procedimiento adecuado. Este es el principal argumento en Hart (1995) a favor de la β-convergencia, pero es importante recordar que condicionar en el pasado o en el futuro es igualmente válido desde un punto de vista estadístico y que las ecuaciones (14), β-convergencia, o (46), γ-convergencia, son igualmente aceptables estadísticamente y contienen la misma información. El futuro puede ser perfectamente exógeno respecto al pasado en un sentido estadístico (Engle, Hendry y Richard (1983)).

En términos de la regresión entre alturas de padres e hijos de Galton (1886a) la ecuación (28) equivale a aquella en la que las alturas de los hijos se explican a partir de las alturas de los padres, esta fue en la que se centró Galton (1886a), y la ecuación (48) aquella en la que las alturas de los padres se explican a partir de las alturas de los hijos. Desde un punto de vista estadístico ambas regresiones tienen idéntica justificación y contienen la misma información, si bien resulta obvio que los hijos no pueden afectar a la altura de sus padres. No es por tanto aceptable extraer conclusiones de causalidad basadas únicamente en la recta de regresión.

Así pues el concepto de **g-convergencia** trata de examinar si las **economías finalmente ricas**, con altos niveles de renta *per capita* en términos relativos, **han tendido a crecer más que las economías finalmente pobres**, con bajos niveles de renta *per capita.* Debido a ello podríamos pensar que si en una muestra concreta encontramos γ-convergencia entonces, debe haberse producido un aumento en la dispersión *cross-section* de la renta *per capita,* en otras palabras debe haberse producido un incremento en la desigualdad en la distribución de la renta; de forma que γ-convergencia implica σ-divergencia. Al igual que sucede con el caso de la β-convergencia este razonamiento intuitivo resulta ser incorrecto, pero es posible derivar una relación entre γ-convergencia y σ-convergencia de forma análoga a la que obtuvimos al analizar la relación entre β-convergencia y σ-convergencia.

Ya hemos observado que en el caso estacionario $Var(y_{it}) = Var(y_{i,t-1}) = \sigma^2 \quad \forall t$, en consecuencia

$$\vartheta^2 = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var^2(y_{it})} = \rho'^2 \leq 1 \quad \Rightarrow \quad |\rho'| < 1 \ \text{ si } \ |\vartheta| < 1 \qquad (50)$$

Por lo tanto **un signo positivo en el coeficiente sobre la condición final en la regresión *cross-section* de convergencia no indica un aumento en la dispersión o desigualdad**, ya que como podemos observar en este ejemplo la dispersión permanece constante pero con tal que la correlación entre la situación inicial y final no sea idénticamente igual a 1 obtendremos γ-convergencia[57].

[57] Tampoco obtendríamos γ-convergencia si la correlación entre la situación inicial y final fuera -1.

En el caso general es posible establecer la siguiente **relación entre g-convergencia y s-convergencia** (Hart y Prais (1956), Prais (1958), Hart (1995)). A partir de la definición del coeficiente de correlación, $\vartheta^2 = \dfrac{Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var(y_{i,t-1})}$, y del coeficiente de γ-convergencia en (48), $\rho' = \dfrac{Cov(y_{it}, y_{i,t-1})}{Var(y_{it})}$, es fácil obtener la siguiente relación

$$\vartheta^2 = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var(y_{i,t-1})} = \frac{Var(y_{it}).Cov^2(y_{it}, y_{i,t-1})}{Var^2(y_{it}).Var(y_{i,t-1})} = \frac{Var(y_{it})}{Var(y_{i,t-1})}.\rho'^2 \qquad (51)$$

por tanto,

$$\frac{Var(y_{i,t-1})}{Var(y_{it})} = \frac{\rho'^2}{\vartheta^2} \qquad (52)$$

lo que nos permite obtener la siguiente **tipología** de casos

(1) $\rho'^2 > 1 \Rightarrow Var(y_{it}) < Var(y_{i,t-1})$ **s-convergencia**

(2) $\vartheta^2 < \rho'^2 < 1 \Rightarrow Var(y_{it}) < Var(y_{i,t-1})$ **s-convergencia**

(3) $\rho'^2 < \vartheta^2 < 1 \Rightarrow Var(y_{it}) > Var(y_{i,t-1})$ **s-divergencia**

En consecuencia **la condición** $\rho'^2 < 1$, o alternativamente $0 < \gamma < 2$, **por sí sola no nos permite alcanzar ninguna conclusión acerca de la evolución en la dispersión de la distribución *cross-section*, aunque** $\rho'^2 > 1$ **si indica de forma inequívoca s-convergencia**.

Obsérvese que ϑ es la **correlación entre la situación inicial y final**, $(y_{i,t-1}, y_{it})$, un estadístico que, como ya hemos mencionado, puede ser utilizado como medida de movilidad intradistribucional, un **valor bajo** de ϑ implica **poca correlación entre la situación inicial y la final**, y por tanto tiende a indicar una **elevada movilidad** dentro de la distribución de $\log x_{it}$ a lo largo del tiempo, por el contrario un **valor alto** de ϑ implica una **elevada correlación entre la**

situación inicial y la final, y por tanto tiende a indicar una **baja movilidad** dentro de la distribución de $\log x_{it}$ a lo largo del tiempo. **Si la movilidad es suficientemente baja**, de forma que $\rho'^2 < \vartheta^2 < 1$, **entonces obtenemos s-divergencia**, $Var(y_{it}) > Var(y_{i,t-1})$.

En **conclusión**, **cuando** $\rho'^2 < 1$ **si la movilidad entre economías**, u observaciones en general, **es baja en relación a la magnitud de la relación entre crecimiento y condición final**, las economías finalmente ricas creciendo más que las finalmente pobres, **entonces la dispersión aumenta y encontraremos una situación de s-divergencia**, pero **por el contrario si la movilidad dentro de la distribución es alta en relación a la magnitud de la relación entre crecimiento y condición final entonces la dispersión se reduce y obtendremos s-convergencia**.

De esta forma para demostrar que economías ricas y pobres están σ-divergiendo debemos condicionar en el periodo inicial y mostrar que $\rho^2 > 1$, mientras que para demostrar que economías ricas y pobres están σ-convergiendo debemos condicionar en el periodo final y mostrar que $\rho'^2 > 1$. El concepto de γ-convergencia es menos popular en teoría del crecimiento que el de β-convergencia porque es menos intuitivo pero ambos contienen la misma información.

Observando que $\rho\rho' = \vartheta^2$ podemos obtener un **resultado** interesante **adicional**,

$$\frac{Var(y_{i,t-1})}{Var(y_{it})} = \frac{\rho'}{\rho}$$

en consecuencia

$$\rho' \lesseqgtr \rho \quad \Leftrightarrow \quad Var(y_{i,t-1}) \lesseqgtr Var(y_{it})$$

por tanto σ-convergencia requiere que $\rho' > \rho$ y σ-divergencia la condición contraria, $\rho' < \rho$.

Finalmente obsérvese que al igual que sucede en el caso de la β-convergencia, el anterior argumento acerca de resultados contrarios entre γ-convergencia y σ-divergencia requiere cierto grado de aleatoriedad entre las situaciones inicial y final, es decir exige que $\vartheta^2 < 1$, cuando dicha relación es de carácter determinista entonces $\vartheta^2 = 1$ y no existe conflicto entre γ-convergencia y σ-divergencia, en el sentido de que $\rho'^2 > 1$, $\gamma < 0$ ó $\gamma > 2$, indica σ-convergencia y al mismo tiempo $\rho'^2 < 1$, $0 < \gamma < 2$, indica inequívocamente σ-divergencia (Prais (1958), p.-269). En este caso $\rho = \frac{1}{\rho'}$, de forma que β-convergencia va necesariamente asociada a γ-divergencia y β-divergencia a γ-convergencia.

b*-convergencia: Datos de panel.*

Excepto por los comentarios realizados al principio de esta sección acerca del concepto de β-convergencia en un contexto de series temporales y el contraste de raíces unidad con datos de panel, toda nuestra exposición en esta sección se ha centrado en la comparación de dos observaciones temporales, ignorando, de esta forma la estructura de *data field* de nuestro conjunto de datos. Hemos pues ignorando información en el análisis y en consecuencia reducido la eficiencia de nuestras estimaciones.

Cuando la muestra se extiende más allá de dos peridos temporales es natural utilizar todo el conjunto de datos disponibles para obtener las mejores estimaciones posibles. Este argumento estuvo ya presente en las aplicaciones iniciales de la ecuación de convergencia, de esta forma Barro y Sala-i-Martín (1991, 1992, 1995) combinaron diferentes subperiodos en la estimación de una única velocidad de convergencia, aunque su método de estimación, referido como **mínimos cuadrados ponderados**, no deja excesivamente claro como se efectúa la estimación y que tipo de efectos fijos, individuales y/o temporales, se están introduciendo.

En el contexto de una muestra que se mueven en dos direcciones, la dimensión temporal, T, y la dimensión *cross-section*, n, un gran conjunto de estimadores están disponibles. Así cuando una dimensión es relativamente reducida en relación a la otra siempre es posible reformular el problema en términos de un **sistema de ecuaciones aparentemente no relacionadas** (**SURE**) y de esta forma introducir cierto grado de **heterogeneidad** en el análisis. En este contexto es fácil la consideración de diversos estimadores de **mínimos cuadrados generalizados** que permitan heterocedasticidad y/o correlación entre las diversas ecuaciones del sistema (Swamy (1971)). Por ejemplo, cuando n es grange en relación a T siempre es posible considerar que disponemos de un sistema de T ecuaciones, estimar parámetros diferentes por periodos, heterogeneidad en la dimensión temporal, y además incorporar heterocedasticidad y/o correlación temporal entre las T ecuaciones del sistema (Barro y Lee (1994a,b), Barro (1999)). Por el contrario, cuando T es grande en relación a n siempre es posible considerar que disponemos de un sistema de n ecuaciones, estimar parámetros diferentes por economías (individuos), heterogeneidad en la dimensión *cross-section*, y además incorporar heterocedasticidad y/o correlación contemporénea entre las economías (individuos), es decir, entre las n ecuaciones del sistema. Este tipo de estimadores no serán, sin embargo, explorados en este trabajo.

Cuando, como en nuestro caso, tanto la dimensión temporal, T, como la dimensión *cross-section*, n, son relativamente grandes o de magnitud similar (*data field*) otras técnicas deben ser consideradas (Quah y Sargent (1993), Quah (1994a), Pesaran y Smith (1995)). En el contexto de los modelos de regresión y dada la estructura de nuestro problema, que podemos considerar representado por la ecuación (27) y que es la misma ecuación en las dos dimensiones de interés, lo natural es la consideración de **técnicas de datos de panel**, que permiten igualmente la introducción de **heterogeneidad** y diversas estructuras de **correlación individual y/o temporal** (Balestra (1992a)). De hecho la estimación de la ecuación de convergencia mediante técnicas de datos de panel, acomodando heterogeneidad inobservable entre economías, ha ganado popularidad en los estudios recientes sobre convergencia económica (Knight, Loayza y Villanueva (1993), Loayza (1994), Canova y Marcet (1995), Islam (1995, 1998), Caselli, Esquivel y Lefort (1996), Boscá (1996), Benhabib y Spiegel (1997), Lee, Pesaran y Smith (1997, 1998), De la Fuente (1998b), Forbes (1998), Gaulier, Hurlin y Jean-Piere (1999), Maddala (1999), Paci y Pigliaru (2000)), y ello a pesar de las críticas de algunos autores (Durlauf y Quah (1998), Section 5).

Por esta razón finalizaremos esta sección examinando algunas de las **ventajas e inconvenientes** de aprovechar la **estructura de panel de las observaciones** en la estimación de (27), así como la forma correcta en que debemos efectuar dicha estimación.

La forma más simple de aprovechar la estructura de panel de la muestra consiste simplemente en apilar las observaciones y estimar la ecuación (27) para el conjunto de $n.T$ observaciones por **mínimos cuadrados ordinarios**, de forma similar a como ya hicimos en el ejercicio de análisis de varianza de la sección anterior. Por las mismas razones que ya aparecieron anteriormente en la realización del contraste de raíces unidad es necesario ampliar dicha ecuación al menos en dos direcciones.

En primer lugar es necesario recoger el **crecimiento** en x_t. Aunque no existe una forma única de hacerlo encontramos que la incorporación en (27) de **efectos fijos temporales** es suficientemente flexible para nuestros propósitos. Estos efectos puede considerarse que **capturan *shocks* globales a la función de producción agregada que son comunes a todas las economías**.

En segundo lugar parece restrictivo suponer que todas las economías son completamente homogéneas y en consecuencia tienen el mismo estado estacionario, por esta razón introducimos en (27) **efectos fijos individuales** destinados a **recoger** la **heterogeneidad inobservable** entre economías[58]. La interpretación natural de estos efectos es considerar que **capturan diferencias peculiares entre economías en los parámetros que caracterizan la función de producción agregada** (Islam (1995))[59], o alternativamente ***shocks* específicos para cada economía**.

De hecho uno de los grandes **atractivos** de aprovechar la estructura de panel de los datos consiste precisamente en permitir dicha **heterogeneidad** sin ser específicos acerca de ella.

[58] La posibilidad de recoger heterogeneidad observable por medio de otras variables explicativas será analizada en Goerlich (2001b).

[59] Desde el punto de vista teórico es posible relacionar dichos efectos fijos con un índice que puede ser interpretado en términos de eficiencia y por tanto constituyen una fuente complementaria de información respecto a la proporcionada por la Productividad Total de los Factores (Islam (1995), Sec.-VIII).

Algunos autores (Islam (1995), Caselli, Esquivel y Lefort (1996, Sec.- 2.2.1)) han argumentado que si dicha heterogeneidad está realmente presente en los datos, lo que es bastante probable en las ecuaciones de convergencia estimadas habitualmente en la literatura, entonces la estimación por mínimos cuadrados ordinarios de ecuaciones de convergencia *cross-section* produce resultados inconsistentes debido a la correlación existente entre los efectos fijos individuales y la condición inicial, log $x_{i,t-j}$. De esta forma un tratamiento consistente de la heterogeneidad sólo es posible si tenemos en cuenta la dimensión temporal de los datos (Pesaran y Smith (1995)).

Al mismo tiempo esta flexibilidad en el tratamiento de la heterogeneidad entre economías constituye un **inconveniente** ya que el aumento en la capacidad explicativa normalmente asociado a la introducción de efectos fijos individuales va acompañado de un abandono en la posibilidad de examinar las causas económicas subyacentes a dicha heterogeneidad. En este sentido **si encontramos β-convergencia en presencia de efectos fijos individuales significativos en realidad estamos argumentando a favor de la existencia de β-convergencia de cada economía a un estado estacionario diferente, lo que ciertamente desvirtúa el concepto de β-convergencia como *catching-up*** dado al comienzo de esta sección (Islam (1995, p.-1162), Durlauf y Quah (1998, p.-50)). En consecuencia deberemos tener presente que la introducción de efectos fijos individuales en la ecuación de convergencia presenta ventajas en términos de flexibilidad, posibilidad de un tratamiento consistente de la heterogeneidad, bondad del ajuste y capacidad explicativa de nuestra ecuación pero también presenta inconvenientes en términos de la capacidad interpretativa de los coeficientes en términos de un modelo teórico. En cualquier caso una vez estimada la ecuación general con efectos fijos incluidos siempre es posible contrastar estadísticamente su significación.

La consideración de efectos fijos, de forma similar a lo que sucede cuando examinamos en concepto de convergencia condicional, introduce cierta heterogeneidad en el análisis y en consecuencia desvirtúa el concepto de β-convergencia. Probablemente es poco informativo encontrar que las economías convergen, pero la convergencia es a diferentes lugares (Islam (1995), p.-1162). **Heterogeneidad e interpretabilidad del concepto de β-convergencia**

presentan un *trade-off* que cualquier analista deberá ponderar adecuadamente en presencia de una muestra concreta.

En definitiva nuestra ecuación de referencia en un contexto de datos de panel viene dada por

$$\log x_{it} - \log x_{i,t-1} = \alpha + \lambda_i + \eta_t - \beta \log x_{i,t-1} + u_{it} \qquad (53)$$

donde λ_i y η_t representan los efectos fijos individuales y temporales, tienen la misma interpretación que en (3) y al igual que en esta ecuación introducimos $\Sigma_{i=1}^{n}\lambda_i = 0$ y $\Sigma_{t=1}^{T-1}\eta_t = 0$ como restricciones de identificación.

Antes de proseguir conviene realizar **tres observaciones** de interés.

- **Primero**, desde el punto de vista de la **dinámica** ya hemos observado como la ecuación (53) carece de generalidad, sólo incluye un desfase.
  Esta cuestión, que ya fue mencionada en el contexto de la interpretación de series temporales del concepto de β-convergencia y en la aplicación de los contrastes de raíces unidad, no parece haber sido incorporada en los trabajos aplicados de estimación de ecuaciones de convergencia con datos de panel (Knight, Loayza y Villanueva (1993), Islam (1995), Caselli, Esquivel y Lefort (1996), De la Fuente (1998b)) y sólo parece haber sido objeto de atención por parte de los trabajos realizados con un enfoque más bien se series temporales (Evans y Karras (1996a, b), Evans (1997), Gaulier, Hurlin y Jean-Piere (1999)).
  Esta puntualización es sin embargo importante, ya que como ha señalado repetidamente la literatura econométrica sobre series temporales los errores de especificación dinámica pueden originar consecuencias graves en lo que a las propiedades de los estimadores se refiere (Hendry (1995)).

- **Segundo**, algunos autores han eliminado en las ecuaciones de convergencia a estimar los efectos fijos temporales mediante la consideración de la variable $z_{it} = \frac{x_{it}}{\mu_t}$ en lugar de x_{it} (Raymond y García-Greciano (1994), De la Fuente (1998b)). De esta forma la ecuación (53) se convierte en

$$\log z_{it} - \log z_{i,t-1} = \alpha + \lambda_i - \beta \log z_{i,t-1} + u_{it} \tag{54}$$

  Puesto que z_{it} no es más que la renta *per capita* respecto al valor del agregado en cada año, esta es una forma adecuada e intuitiva de **eliminar tendencias** en los datos **y** por tanto también **los *shocks* globales a la función de producción agregada que son comunes a todas las economías** y que eran captados por los efectos fijos temporales[60]. En consecuencia la consideración de z_{it} en lugar de x_{it} es igualmente adecuada en el contexto de la ecuación de convergencia, aunque debe observarse que (53) y (54) no son numéricamente equivalentes, salvo en el caso improbable en que

$$\eta_t = \log \mu_t - (1-\beta) \log \mu_{t-1}$$

  Que ecuación, (53) ó (54), es más adecuada es una cuestión de bondad del ajuste en cada caso particular.

- **Tercero**, **heterogeneidad** a través de efectos fijos individuales, λ_i en (53), será la única que consideraremos en este trabajo, de forma que las economías convergen a diferentes niveles de renta *per capita* en el estado estacionario, pero la tasa de crecimiento en dicho estado y la velocidad de convergencia hacia el mismo es la misma para todos los países. Un mayor grado de heterogeneidad es posible, pero como ya hemos argumentado anteriormente a mayor heterogeneidad menor interpretabilidad del concepto de β-convergencia.

[60] Puesto que la media de agregado en cada año, μ_t, es una media ponderada y la regresión (54) no utiliza ponderaciones la constante, α, no puede ser eliminada ya que las variables incluidas en dicha ecuación no tienen media simple cero (De la Fuente (1998a), p.-2).

La introducción de **variables explicativas adicionales**, que difieran entre países y en el tiempo, en la ecuación de convergencia (**convergencia condicional**) será considerada en Goerlich (2001b). Mayor heterogeneidad todavía es posible, de hecho algunos autores han argumentado que la heterogeneidad en las muestras utilizadas habitualmente en el análisis aplicado sobre el crecimiento económico es considerable y puede distorsionar gravemente los resultados (Lee, Pesaran y Smith (1995, 1997, 1998)) de forma que lo esencial es que pueden decirnos los métods econométricos sobre valores medios de los parámetros y en este sentido han propuesto estimar ecuaciones de convergencia para cada individuo y construir a partir de las estimaciones individuales **estimadores de medias de grupos** (Pesaran y Smith (1995), Boscá (1996)) que tomen en cuenta toda la heterogeneidad existente en la muestra[61]. Aunque estos estimadores son factibles su interpretación y utilidad en términos teóricos es dudosa (Islam (1998)) y no serán considerados.

El estudio de la heterogeneidad es sin embargo importante (Durlauf y Johnson (1995), Hansen (2000)) y en este sentido los argumentos en Pesaran y Smith (1995) sugieren que en presencia de elevada heterogeneidad la regresión *cross-section* basada en largas medias temporales puede ser preferible a la imposición de coeficientes homogéneos en un contexto de datos de panel (*pooling*), en el sentido de que la regresión *cross-section* produce estimaciones consistentes de los coeficientes medios de largo plazo mientras que el *pooling* no. La utilidad de estos argumentos depende probablemente de cada muestra concreta y resulta difícil realizar afirmaciones con carácter general.

Dado que en la práctica más habitual el espaciado temporal entre observaciones no suele ser de un solo periodo la ecuación que generalmente se estima es la análoga a (36) para datos de panel, es decir

$$\frac{1}{j}\left(\log x_{it} - \log x_{i,t-j}\right) = a + \lambda_i + \eta_t - b' \log x_{i,t-j} + \varepsilon_{it} \tag{55}$$

donde los términos λ_i y η_t son idénticos a los de (53).

---

[61] El procedimiento es similar al propuesto para el contrastde de raíces unidad en el contexto de paneles heterogéneos (Im, Pesaran y Shin (1997), Maddala y Wu (1999)).

La determinación del **valor apropiado de *j*** ha sido objeto de alguna breve mención en la literatura, de hecho el paso de una regresión *cross-section* a la utilización de datos de panel implica normalmente la subdivisión de un periodo temporal largo en subperiodos más pequeños, pero ¿como de pequeños deben ser estos subperiodos?. **¿Sobre que diménsión temporal se supone que la ecuación de convergencia es válida?**. Aquí encontramos de nuevo un *trade-off* al que habrá que hacer frente en términos prácticos. Desde un **argumento** estrictamente **estadístico** cuantos más datos mejor, lo que apunta hacia **subperiodos lo más pequeños posibles**, de hecho en términos teóricos la ecuación de convergencia (33) representa una aproximación al estado estacionario igualmente válida sobre periodos cortos de tiempo que sobre periodos largos (Islam (1995), p.-1137); pero **¿significa esto que podemos utilizar el modelo de crecimiento neoclásico para explicar la dinámica semanal o incluso diaria de la renta *per capita* al mismo tiempo que los movimientos de décadas o incluso siglos?**, ciertamente no. Diferentes lapsos temporales para analizar el proceso de crecimiento sólo serán igualmente válidos si los problemas de especificación del modelo son independientes de su escala temporal, lo que no es probable que suceda en el caso de la teoría del crecimiento, donde los problemas de especificación son con toda probabilidad mayores en el corto plazo (Durlauf y Quah (1998), p.-51). **La consideración de periodos excesivamente cortos**, digamos anuales, **introducen en el análisis perturbaciones cíclicas que oscurecen la dinámica de largo plazo**. Por esta razón la literatura ha considerado generalmente, sin ningún tipo de justificación teórica, lapsos temporales entre 5 y 10 años (Barro y Sala-i-Martín (1992), Islam (1995), Caselli, Esquivel y Lefort (1996)), aunque aplicaciones con lapsos temporales inferiores también son frecuentes (Lee, Pesaran y Smith (1997, 1998), De la Fuente (1998b)).

Desde el punto de vista meramente práctico los trabajos que utilizan técnicas de panel tienden a obtener velocidades de convergencia sensiblemente superiores a las obtenidas a partir de ecuaciones *cross-section*. Así, Islam (1995) obtiene velocidades entre el 3.8% y el 9.1%, Caselli, Esquivel y Lefort (1996) en el entorno de 10%, Boscá (1996) valores entre el 11% y el 39%, Lee, Pesaran y Smith (1997) aproximadamente el 30%, aunque estos últimos trabajos permiten un mayor grado de heterogeneidad, y De la Fuente (1998b) alrededor del 8% anual.

Aumentar la frecuencia de los datos, aumenta pues **la velocidad de convergencia** y en consecuencia **disminuye la persistencia** respecto a las estimaciones *cross-section*.

La ecuación (55) nos permite además calcular las **rentas *per capita* de estado estacionario** correspondientes a cada economía en ausencia de efectos fijos temporales, que por su propia naturaleza no pueden ser extrapolados. A partir de la versión determinista de (55) obtenemos que, en el estado estacionario, la renta *per capita* de la economía *i*, x_i^*, puede obtenerse como

$$x_i^* = \exp\left(\frac{a+\lambda_i}{b'}\right)$$

A título de ejemplo si consideramos la renta per capita provincial en subperiodos decenales para el periodo 1955-1995, la estimación de (55) **mínimos cuadrados ordinarios** generó el siguiente resultado, $\hat{b}' = 0.0367$, lo que representa una **velocidad de convergencia** anual del $\hat{b} = 4.57\%$, altamente significativa a juzgar por los métodos convencionales, y un $R^2 = 89.87\%$. En consecuencia la adición de la condición inicial al análisis de varianza considerado en la sección anterior mejora sensiblemente, en algo más de tres puntos porcentuales, la capacidad explicativa de nuestro modelo.

El principal **problema** con la estimación de la **ecuación (55)** por **mínimos cuadrados ordinarios** y datos de panel es que dicha ecuación es un modelo dinámico. Ello se observa más claramente si escribimos la ecuación (53) en forma autoregresiva

$$\log x_{it} = \alpha + \lambda_i + \eta_t + \rho \log x_{i,t-1} + u_{it} \qquad \rho = 1 - \beta \tag{56}$$

cuya única diferencia con (28) consiste en la presencia de efectos fijos individuales y temporales aprovechando la estructura de panel de los datos.

Es bien conocido en la literatura econométrica que la presencia de **efectos fijos**[62] **individuales** en **modelos dinámicos** de **datos de panel** provoca que el **estimador** de **mínimos cuadrados ordinarios** en (56) sea **inconsistente** cuando $n \to \infty$ y ***T* es fijo** (Nickell (1981), Sevestre y Trognon (1985, 1992), Hsiao (1986, Sec.-4.2)). La razón estriba en que aún suponiendo que u_{it} en (56) sea independiente e idénticamente distribuido, tanto en el corte transversal como en la dimensión temporal, la inclusión de los términos $\alpha + \lambda_i$ en esta ecuación es equivalente a estimar el modelo en desviaciones respecto las medias individuales (Frisch y Waugh (1933)) y aunque $y_{i,t-1}$ y u_{it} no estén correlacionados, sus respectivas medias individuales, $\bar{y}_{i,\bullet-1} = \frac{1}{T}\Sigma_{t=2}^{T} y_{i,t-1}$ y $\bar{u}_{i,\bullet} = \frac{1}{T}\Sigma_{t=2}^{T} u_{it}$, si lo están, (*i*) entre ellas, (*ii*) $\bar{y}_{i,\bullet-1}$ con u_{it} y (*iii*) $\bar{u}_{i,\bullet}$ con $y_{i,t-1}$, y la suma de estos tres términos de covarianza no desaparece. Sin embargo si consideramos el caso en el que $T \to \infty$ entonces el estimador de mínimos cuadrados ordinarios en (56) si es consistente y asintóticamente equivalente al estimador de máxima verosimilitud bajo normalidad[63] (Amemiya (1967)).

Sabemos además que cuando T es pequeño y $\rho > 0$ entonces el sesgo de mínimos cuadrados ordinarios es negativo (Hsiao (1986), Sec.-4.2). En consecuencia, puesto que $\beta = 1 - \rho$, es probable que en estas situaciones las **estimaciones de la velocidad de convergencia derivadas de (54) y (55) estén sesgadas al alza**. No obstante aunque algunos autores opinan que la utilización de métodos de estimación especificamente diseñados para paneles dinámicos introduce grados de sofisticación innecesarios en el análisis (Temple (1999), p.-132) es cierto que parte de la discusión reciente sobre la cuestión de la convergencia económica se ha centrado en torno a la adecuada utilización de los métodos econométricos en estos casos (Islam (1995), Caselli, Esquivel y Lefort (1996), Forbes (1998), Aghion, Caroli y García-Peñalosa (1999, p.-1618)) y ello dejando al margen el tema ya mencionado de la heterogeneidad.

[62] La constante, α, es suficiente para generar el resultado que mencionaremos a continuación.
[63] Este resultado es similar al caso de un AR(1) en series temporales. En este caso sabemos que el estimador de mínimos cuadrados ordinarios es sesgado en muestras finitas pero consistente conforme $T \to \infty$. Con datos de panel el estimador de mínimos cuadrados ordinarios no sólo es sesgado en muestras finitas sino también inconsistente mientras que T sea fijo y esta inconsistencia sólo desaparece cuando $T \to \infty$.

La estimación consistente de (56), o alternativamente (53), cuando T es fijo ha sido objeto de atención teórica por parte de la literatura econométrica desde el trabajo original de Balestra y Nerlove (1966) y disponemos en la actualidad de una serie de estimadores consistentes bajo determinados supuestos y con diferentes grados de eficiencia (Anderson y Hsiao (1981, 1982), Chamberlain (1982, 1984), Bhargava y Sargan (1983), Holtz-Eakin, Newey y Rosen (1988), Arellano (1989), Arellano y Bond (1991), Keane y Runkle (1992), Ahn y Schmidt (1995), Arellano y Bover (1995), Blundell y Bond (1998)). La literatura no parece haber alcanzado, sin embargo, un consenso unánime sobre el estimador más adecuado con generalidad, si bien los últimos trabajos al respecto parecen haber avanzado considerablemente en la cuestión (Blundell y Bond (1998)).

Puesto que el problema en la ecuación (56) lo generan los efectos fijos una solución que se ha mostrado útil en la práctica consiste en eliminarlos mediante **diferenciación temporal** (Arellano and Bond (1991)). Diferenciando esta ecuación

$$\log x_{it} - \log x_{i,t-1} = (\eta_t - \eta_{t-1}) + \rho(\log x_{i,t-1} - \log x_{i,t-2}) + (u_{it} - u_{i,t-1}) \qquad (57)$$

o

$$\Delta \log x_{it} = \Delta\eta_t + \rho\Delta \log x_{i,t-1} + \Delta u_{it} \qquad (58)$$

donde Δ es el operador diferencia temporal, $\Delta \log x_{it} = \log x_{it} - \log x_{i,t-1}$.

En términos de β (58) puede ser escrito como

$$\Delta \log x_{it} - \Delta \log x_{i,t-1} = \Delta\eta_t - \beta\Delta \log x_{i,t-1} + \Delta u_{it} \qquad (59)$$

y dado que el lapso temporal entre observaciones no suele ser igual a un periodo la ecuación que se estima es análoga a (55)

$$\frac{1}{j}\left(\Delta \log x_{it} - \Delta \log x_{i,t-j}\right) = \Delta\eta_t - b'\Delta \log x_{i,t-j} + \Delta\varepsilon_{it} \qquad (60)$$

donde las propiedades de $\Delta\varepsilon_{it}$ son idénticas a las de Δu_{it}, excepto por un factor de proporcionalidad, $\frac{1}{j}$.

En términos del modelo neoclásico de crecimiento la ecuación (60) implica una relación entre la aceleración del crecimiento y la tasa de crecimiento inicial de la renta *per capita*, de forma que ahora tratamos de explicar la segunda derivada a partir de la primera.

La diferenciación sin embargo crea un nuevo problema, $\log x_{i,t-1}$ está correlacionado con el término de perturbación a través de $u_{i,t-1}$, y por tanto la **estimación de (57) por mínimos cuadrados ordinarios no es apropiada**[64]. Sin embargo $\log x_{i,t-2}$ no está correlacionado con el término de perturbación $(u_{it} - u_{i,t-1})$ y en consecuencia puede ser utilizado como instrumento en la estimación de (57) por **variables instrumentales** (Anderson and Hsiao (1981)). Es más, $\log x_{i,t-s}$ $s \geq 2$ no presenta correlación con $(u_{it} - u_{i,t-1})$ con lo que obtenemos para $T \geq 3$ los siguientes momentos poblacionales

$$\mathrm{E}\left[\left(\Delta \log x_{it} - \Delta\eta_t - \rho\Delta \log x_{i,t-1}\right)\log x_{i,t-s}\right] = 0 \qquad s = 2,\ldots,(t-1);\ t = 3,\ldots,T \qquad (61)$$

que en total representan $m = (T-2).(T-1)/2$ restriciones lineales que pueden ser utilizadas en la estimación.

Considerando a (57) como un sistema hipotético de $T-2$ ecuaciones podemos construir a partir de (61) un **estimador generalizado de momentos** de ρ que sea óptimo dentro de su

[64] Obsérvese que esta correlación es fruto de los supuestos acerca de u_{it} en la ecuación de partida. Si hubiéramos supuesto que u_{it} en la ecuación en niveles es un paseo aleatorio con pertubaciones i.i.d. entonces no se produciría esta correlación y la estimación de (57) por mínimos cuadrados ordinarios sería adecuada (Arellano y Bond (1991), p.-282).

clase cuando $n \to \infty$ y T es fijo en la linea sugerida por Hansen (1982) y White (1982) y en el que el número de instrumentos varía de ecuación a ecuación, incrementándose con t (Arellano y Bond (1991)). En concreto, al margen de los efectos fijos temporales, obtenemos la siguiente **asignación dinámica de instrumentos** para cada ecuación

$$
\begin{array}{cccc}
t=3 & t=4 & \cdots\cdots & t=T \\
\begin{bmatrix} \log x_{11} \\ \log x_{21} \\ \vdots \\ \log x_{n1} \end{bmatrix} &
\begin{bmatrix} \log x_{11} & \log x_{12} \\ \log x_{21} & \log x_{22} \\ \vdots & \vdots \\ \log x_{n1} & \log x_{n2} \end{bmatrix} &
\cdots\cdots &
\begin{bmatrix} \log x_{11} & \log x_{12} & \cdots & \log x_{1,T-2} \\ \log x_{21} & \log x_{22} & \cdots & \log x_{2,T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \log x_{n1} & \log x_{n2} & \cdots & \log x_{n,T-2} \end{bmatrix}
\end{array}
$$

De esta forma utilizamos **instrumentos desfasados en niveles para estimar una ecuación en diferencias**. Todo ello en ausencia de información adicional acerca de las condiciones iniciales del proceso y simplemente suponiendo que u_{it} es independiente e idénticamente distribuido tanto en el corte transversal como en la dimensión temporal. Son precisamente las restricciones sobre la matriz de covarianzas del término de perturbación en (56) las que permiten la estimación de los parámetros de interés (Ahn y Schmidt (1995)). Veremos a continuación como estas condiciones son más fuertes de lo necesario pero facilitan la derivación e intuición del estimador.

Dadas las propiedades i.i.d. supuestas para u_{it} obtenemos que

$$(i) \qquad \mathrm{Var}(u_{it} - u_{i,t-1}) = 2\sigma_u^2 \qquad \forall i,t$$

$$(ii) \qquad \mathrm{Cov}\left[(u_{it} - u_{i,t-1}).(u_{i,t-1} - u_{i,t-2})\right] = -\sigma_u^2 \qquad \forall i,t$$

$$(iii) \qquad \mathrm{Cov}\left[(u_{it} - u_{i,t-1}).(u_{i,t-s} - u_{i,t-s-1})\right] = 0 \qquad \forall i,t,s>1$$

$$(iv) \qquad \mathrm{Cov}\left[(u_{it} - u_{i,t-1}).(u_{j,t-s} - u_{j,t-s-1})\right] = 0 \qquad \forall i \neq j, t,s$$

en consecuencia la matriz de varianzas-covarianzas del término de perturbación en (57) es conocida hasta un factor de proporcionalidad.

De la misma forma que hicimos con el modelo (3) de la sección anterior podemos formular el modelo (56) en **notación de muestra completa** suponiendo una **organización de las observaciones por individuo**. Teniendo en cuenta que el carácter dinámico de (56) nos hace perder una observación temporal para cada individuo podemos escribir

$$\log \mathbf{x} = \ell_{n(T-1)}\alpha + \mathbf{D}_n \mathbf{l} + \mathbf{D}_{T-1}\mathbf{h} + \log \mathbf{x}_{-1}\rho + \mathbf{u} \qquad (62)$$

donde $\log \mathbf{x}$ es el vector $n(T-1)$x1 de los logaritmos de las rentas *per capita*, $\ell_{n(T-1)}$ es un vector de unos de dimensión $n(T-1)$, $\mathbf{D}_n = \mathbf{I}_n \otimes \ell_{T-1}$ es una matriz $n(T-1)$xn que contiene el conjunto de las n variables ficticias individuales, $\mathbf{l}$ es un vector nx1 de efectos fijos individuales, $\mathbf{D}_{T-1} = \ell_n \otimes \mathbf{I}_{T-1}$ es una matriz $n(T-1)$x$(T-1)$ que contiene el conjunto de las $T-1$ variables ficticias temporales, $\mathbf{h}$ es un vector $(T-1)$x1 de efectos fijos temporales y $\mathbf{u}$ es el vector $n(T-1)$x1 de perturbaciones.

Definiendo la matriz de diferenciación temporal para cada individuo, *i*

$$\mathbf{A}_\Delta = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}_{(T-2)\text{x}(T-1)}$$

y para todo los individuos como

$$\mathbf{A} = \mathbf{I}_n \otimes \mathbf{A}_\Delta = \begin{bmatrix} \overbrace{\mathbf{A}_\Delta \quad \cdots \quad \mathbf{0}}^{n} \\ \vdots \qquad \ddots \qquad \vdots \\ \mathbf{0} \quad \cdots \quad \mathbf{A}_\Delta \end{bmatrix}_{n(T-2)\text{x}n(T-1)}$$

podemos escribir el modelo (57) en notación de muestra completa premultiplicando (62) por **A**.

Observando que $\mathbf{A}_{\Delta}\ell_{T-1} = \mathbf{0}$ implica

$$(i) \qquad \mathbf{A}\ell_{n(T-1)} = \mathbf{0}$$

y

$$(ii) \qquad \mathbf{AD}_n = (\mathbf{I}_n \otimes \mathbf{A}_{\Delta})(\mathbf{I}_n \otimes \ell_{T-1}) = \mathbf{0}$$

entonces obtenemos que el hipotético conjunto de $T-2$ ecuaciones representado por (57) puede ser expresado en notación de muestra completa como

$$\mathbf{A}\log\mathbf{x} = \mathbf{AD}_{T-1}\boldsymbol{\eta} + \mathbf{A}\log\mathbf{x}_{-1}\rho + \mathbf{Au} \tag{63}$$

Como ya mencionamos anteriormente esta ecuación será estimada sujeta a la restricción de identificación $\ell'_{T-1}\eta = 0$. Obsérvese que esta restricción no es ahora estrictamente necesaria ya que la constante, α, y los efectos fijos, λ_i, han sido eliminados. Su incorporación se hace simplemente para mantener la equivalencia entre la ecuación en diferencias y la ecuación en niveles de partida. En la práctica, sin embargo, puede ser más conveniente o simplemente más sencillo no introducirla, al margen de que existan otros procedimientos operativos para tomar en consideración los efectos fijos temporales. Por ejemplo podríamos simplemente redefinir $T-2$ efectos fijos en (63) e introducirlos sin restringir, como parece hacerse en las aplicaciones prácticas (Arellano y Bond (1991, Sec.-5), Blundell y Bond (1998, Sec.-7)), considerar las variables en desviaciones respecto a las medias en cada periodo para evitar la necesidad de incluir los efectos fijos temporales en la estimación al eliminar este procedimiento los términos η_t en (56) (Caselli, Esquivel y Lefort (1996)), o simplemente centrar nuestro análisis empírico en $z_{it} = \dfrac{x_{it}}{\mu_t}$ en lugar de en la variable x_{it} (De la Fuente (1998b)). No obstante debe observarse que una estricta equivalencia entre la ecuación en niveles (56) de la que hemos partido y la ecuación transformada (57) requiere la estimación de los $T-1$ efectos fijos temporales originales con la restricción $\ell'_{T-1}\eta = 0$ incorporada en el análisis.

Adicionalmente dadas las propiedades para u_{it} entonces si $\mathbf{u}_i$ es el vector $(T-1)$x1 de perturbaciones en (62) para el individuo i, $\mathbf{u}_i = (u_{i2}, u_{i3}, \ldots, u_{iT})'$, se obtiene que $\mathrm{Var}\left(\mathbf{A}_\Delta \mathbf{u}_i\right) = \sigma_u^2 \mathbf{A}_\Delta \mathbf{A}'_\Delta = \sigma_u^2 \mathbf{H}$, siendo

$$\mathbf{H} = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{bmatrix}_{(T-2)\mathrm{x}(T-2)}$$

por tanto $\mathrm{Var}\left(\mathbf{Au}\right) = \sigma_u^2 (\mathbf{I}_n \otimes \mathbf{H})$.

Así pues dada la matriz de **instrumentos**, **Z**, que estará constituida por:

(i) las **variables ficticias** que recogen los efectos fijos temporales,

$$\mathbf{AD}_{T-1} = (\mathbf{I}_n \otimes \mathbf{A}_\Delta)(\ell_n \otimes \mathbf{I}_{T-1}) = \ell_n \otimes \mathbf{A}_\Delta$$

y que deberá incluir la restricción $\ell'_{T-1}\eta = 0$ en el caso de que esta sea incorporada en la estimación, y

(ii) los **instrumentos válidos** derivados de las restricciones lineales de momentos (61), que para el individuo *i* vienen dados por

$$\mathbf{Z}_i^* = \begin{bmatrix} \log x_{i1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \log x_{i1} & \log x_{i2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \log x_{i1} & \log x_{i2} & \cdots & \log x_{i,T-2} \end{bmatrix}_{(T-2)\mathrm{x}m}$$

siendo $m = (T-2).(T-1)/2$ el número de instrumentos derivados de (61) utilizados en el proceso de estimación. Para el individuo *i* la matriz completa de instrumentos es pues

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{A}_\Delta & \mathbf{Z}_i^* \end{bmatrix}_{(T-2)\mathrm{x}[(T-1)+m]}$$

y para el sistema (63) dicha matriz viene dada por

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_n \end{bmatrix} = \begin{bmatrix} \mathbf{A}_\Delta & \mathbf{Z}_1^* \\ \mathbf{A}_\Delta & \mathbf{Z}_2^* \\ \vdots & \vdots \\ \mathbf{A}_\Delta & \mathbf{Z}_n^* \end{bmatrix} = \begin{bmatrix} \ell_n \otimes \mathbf{A}_\Delta & \mathbf{Z}^* \end{bmatrix}_{n(T-2)\mathrm{x}[(T-1)+m]}$$

De esta forma dado el sistema (63), la matriz de instrumentos $\mathbf{Z}$ y $\mathrm{Var}(\mathbf{Au}) = \sigma_u^2(\mathbf{I}_n \otimes \mathbf{H})$, conocida hasta el factor de proporcionalidad $\mathbf{s}_u^2$, es directo obtener un **estimador generalizado de momentos óptimo** bajo nuestras hipótesis hacerca de u_{it} utilizando como **matriz de ponderaciones** (Hansen (1982), White (1982))

$$\mathbf{W}_n = \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{Z}_i' \mathbf{H} \mathbf{Z}_i \right)^{-1} = \left(\frac{1}{n} \mathbf{Z}' (\mathbf{I}_n \otimes \mathbf{H}) \mathbf{Z} \right)^{-1} \qquad (64)$$

Este **estimador** puede ser calculado en **una sola etapa** y es **óptimo** bajo el **supuesto** de **que** $\boldsymbol{u_{it}}$ **es independiente e idénticamente distribuido**, **tanto en el corte transversal como en la dimensión temporal**, dentro de la clase de estimadores basados en las restricciones de momentos lineales (63).

Si **mantenemos** el supuesto de **independencia**[65] pero **permitimos** la existencia de **heterocedasticidad de forma desconocida**, **tanto en el corte transversal como en la**

[65] En la dimensión temporal ausencia de correlación sería suficiente, aunque deberemos mantener independencia en el corte transversal. Con independencia en las dos dimensiones la matriz de ponderaciones para construir el estimador generalizado de momentos óptimo tomaría una forma algo más sencilla (Arellano y Bond (1991, p.-279)).

dimensión temporal, entonces el estimador que utiliza como matriz de ponderaciones $\mathbf{W}_n$ sería consistente, pero no eficiente. No obstante es siempre posible obtener un **estimador robusto frente a formas arbitrarias de heterocedasticidad** que sea **óptimo** dentro de la clase de estimadores basados en las restricciones de momentos lineales (61) y en ausencia de hipótesis adicionales. Este es necesariamente un **estimador en dos etapas**.

Dados los resíduos de (63) obtenidos a partir de un estimador consistente en una primera etapa, y que como elección natural se obtiene utilizando $\mathbf{W}_n = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{H}\mathbf{Z}_i\right)^{-1}$ como matriz de ponderaciones (Arellano y Bond (1991), p.-279), entonces el **estimador eficiente robusto frente a heterocedasticidad de forma desconocida** se obtiene en una **segunda etapa** utilizando como **matriz de ponderaciones**

$$\widetilde{\mathbf{W}}_n = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i\right)^{-1} \tag{65}$$

donde $\hat{\mathbf{u}}_i = (\hat{\upsilon}_{i3}, \hat{\upsilon}_{i4}, \ldots, \hat{\upsilon}_{iT})'$ es el vector $(T-2)\mathrm{x}1$ de resíduos de primera etapa de (63) para el individuo *i*. En general para todo el sistema el vector de resíduos de primera etapa viene dado por el vector $\hat{\mathbf{u}} = \mathbf{A}\hat{\mathbf{u}} = (\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \ldots, \hat{\mathbf{u}}_n)'$ de dimensión $n(T-2)\mathrm{x}1$.

Como ya hemos indicado el estimador de momentos que utiliza en una **primera etapa** $\mathbf{W}_n$ y en una **segunda etapa** $\widetilde{\mathbf{W}}_n$ es un **estimador generalizado de momentos óptimo**, **bajo independencia en el corte transversal y ausencia de correlación serial en la dimensión temporal, y robusto frente a heterocedasticidad de forma desconocida en ambas direcciones**, **dentro de la clase de estimadores basados en las restricciones de momentos lineales (61) y en ausencia de hipótesis adicionales** (Hansen (1982), White (1982)).

Es importante señalar que puesto que la identificación de los parámetros de interés ha sido posible gracias al supuesto de ausencia de correlación serial en u_{it} conviene contrastar esta hipótesis en la ecuación (63) donde la implicación es que los resíduos diferenciados de esta ecuación no deben mostrar síntomas de correlación de segundo orden. Este es en realidad un contraste sobre la validez de los instrumentos y el contraste de restricciones de sobreidentificación de Sargan (1958, 1988)-Hansen (1982) es apropiado para este fin además de ser capaz de detectar otros posibles problemas de especificación. Existen, no obstante, contrastes específicos para este problema en el contexto de datos de panel (Arellano y Bond (1991), Sec.- 3), o es posible alternativamente construir estadísticos concretos a partir del principio de Hausman (1978).

Finalmente indicar que si los **efectos fijos individuales** fueran de interés una burda estimación puntual de los mismos puede ser obtenida de la siguiente forma. A partir de la estimación de (57) y sustituyendo dichas estimaciones en (56) es posible obtener una estimación de $\alpha + \lambda_i + u_{it}$ como

$$\overline{(\alpha + \lambda_i + u_{it})} = \log x_{it} - (\hat{\eta}_t + \hat{\rho} \log x_{i,t-1}) \tag{66}$$

y promediando estas observaciones en el tiempo es posible obtener una estimación puntual de los efectos fijos individuales que nos dé una idea aproximada de la magnitud de los mismos

$$\widehat{(\alpha + \lambda_i)} = \frac{1}{T-1} \sum_{t=1}^{T-1} \overline{(\alpha + \lambda_i + u_{it})} \tag{67}$$

Antes de finalizar conviene realizar algunas **observaciones** de interés.

- Alternativamente a $\log x_{i,t-2}$ podríamos utilizar como **instrumentos primeras diferencias**, $\Delta \log x_{i,t-2}$ o en general $\Delta \log x_{i,t-s}$ $s \geq 2$ (Anderson y Hsiao (1981)), pero ello haría que nuestro estimador fuera ineficiente y además sólo estaría definido para $T \geq 4$.

- Los supuestos acerca de u_{it} también implican **restricciones de momentos no lineales**, en concreto **cuadráticas** (Arellano y Bond (1991), Ahn y Schmidt (1995)), lo que implicaría un **estimador de momentos generalizado no lineal**.

- Otras transformaciones alternativas a las primeras diferencias temporales, tales como **desviaciones ortogonales** que mantienen la ausencia de autocorrelación en u_{it}, han sido exploradas por la literatura (Arellano (1988), Arellano y Bover (1995)).

- Puesto que la solución apuntada consiste en eliminar los efectos individuales, λ_i en (56), en la práctica suele ser indiferente tratar a estos **efectos** como **fijos o** como **aleatorios**, al menos como punto de partida conceptual. Por esta razón la mayor parte de literatura teórica que estudia el tipo de estimadores que estamos examinando utiliza generalmente el supuesto de que λ_i es un efecto aleatorio y por tanto la ecuación (56) es un modelo con **término de perturbación compuesto** (*error component model*), $\lambda_i + u_{it}$, en el que λ_i se supone, como primera aproximación, independiente e idénticamente distribuido, con $\mathrm{E}(\lambda_i) = 0 \ \forall i$ e independiente de $u_{it} \ \forall i,t$ (Keane y Runkle (1992), Ahn y Schmidt (1995), Arellano y Bover (1995), Blundell y Bond (1998)).

  En principio lo mismo podría hacerse con los efectos temporales, η_t, aunque estos no suelen ser incorporados en el análisis teórico y en las aplicaciones prácticas suelen mantenerse como fijos.

- En cualquier caso seguiremos insistiendo en el carácter de **efecto fijo** de λ_i dado que (*i*) en nuestro contexto la muestra suele ser exahustiva, paises o regiones, y (*ii*) es bastante probable que los efectos individuales estén correlacionados con $\log x_{i,t-1}$. Razones que justifican la especificación de la ecuación de convergencia en términos de efectos fijos (Sevestre y Trognon (1992), p.-97).

- En el contexto de modelos con términos de perturbación compuesto, Ahn y Schmidt (1995) muestran como el supuesto de **homocedasticidad** temporal, que hemos realizado, implica **restricciones de momentos adicionales**, **lineales y no lineales**, que pueden ser potencialmente incorporadas en el proceso de estimación. Estos autores examinan además hasta que punto es posible debilitar los supuestos sobre la perturbación en (56) sin que ello afecte a los momentos utilizados en el proceso de estimación.

- En el mismo contexto, Blundell y Bond (1998) muestran como con una **restricción** adicional **sobre la condición inicial** del proceso generador de $\log x_{i1}$ es posible utilizar un **estimador generalizado de momentos lineal** que utilice todas las restricciones sobre los segundos momentos de la perturbación. En consecuencia los supuestos sobre las condiciones iniciales son importantes en esta clase de modelos, ya que con T finito dichas condiciones afectan a la eficiencia asintótica de los estimadores. **Un estimador más eficiente utilizará no sólo instrumentos desfasados en niveles para una ecuación en diferencias sino también instrumentos desfasados en diferencias para una ecuación en niveles** (Arellano y Bover (1995), Blundell y Bond (1998)).

- Por tanto debemos tener presente que el estimador que hemos presentado es potencialmente mejorable en términos de eficiencia. De hecho existe cierta evidencia de Monte Carlo de que dicho estimador, que utiliza sólo los momentos (61), tiene sesgos apreciables en muestras finitas (Alonso-Borrego y Arellano (1996)) y una cierta tendencia a obtener errores estándar sesgados a la baja (Arellano y Bond (1991), p.-293). La razón de este pobre comportamiento de nuestro estimador de momentos en muestras finitas radica en que para valores de ρ moderadamente grandes y cuando el valor de T es relativamente pequeño, niveles desfasados de la variable suelen ser instrumentos débiles para las primeras diferencias (Blundell y Bond (1998)).

- Los estimadores de **variables instrumentales** han sido utilizados en el contexto de ecuaciones de convergencia como mecanismo para solucionar los problemas derivados de la posible endogenidad de variables explicativas adicionales, es decir en **ecuaciones de convergencia condicionadas** (Barro y Lee (1994a,b), Barro (1999))[66], pero sólo

[66] Para evitar estos problemas es conveniente fechar todas las variables explicativas adicionales a principio del periodo, es decir en el mismo periodo que la condición inicial (Cho (1996)).

recientemente se han introducido en el contexto de paneles dinámicos (Islam (1995), Caselli, Esquivel y Lefort (1996), De la Fuente (1998b), Forbes (1998)).

- Como muestra de la sensibilidad de las estimaciones a los diferentes métodos de estimación baste señalar que cuando estimamos la ecuación (60) para el periodo 1955-1995, subperiodos decenales y utilizando $\mathbf{W}_n$ como matriz de ponderaciones, obtuvimos $\hat{b}' = 0.0131$, lo que implica una velocidad de convergencia anual del $\hat{b} = 1.40\%$. Notablemente más baja que en el caso de la estimación por mínimos cuadrados ordinarios. Además, si utilizanmos $\tilde{\mathbf{W}}_n$ como matriz de ponderaciones en una segunda etapa, la estimación de b' de dos etapas disminuye a aproximadamente la mitad, $\hat{b}' = 0.0067$, lo que representa una velocidad de convergencia anual muy baja, $\hat{b} = 0.69\%$

En **resumen**, como han señalado otros autores, la aplicación práctica del concepto de β-convergencia es ampliamente sensible a los métodos econométricos utilizados, tanto en relación a la consideración de la estructura de panel de los datos como a las implicaciones dinámicas que ello tiene y a la posible heterogeneidad en la muestra. Por ello todavía hay un amplio margen para el debate acerca de cual es la velocidad de convergencia entre regiones o países, al mismo tiempo que la utilidad del propio concepto de β-convergencia se hace cada vez más difusa.

Referencias

Aghion, P.; Caroli, E. & García-Peñalosa, C. (1999) "Inequality and economic growth: The perspective of the new growth theories", *Journal of Economic Literature*, 37, (December), 1615-1660.

Ahn, S. C. & Schmidt, P. (1995) "Efficient estimation of models for dynamic panel data", *Journal of Econometrics*, 68, 5-27.

Alonso-Borrego, C. y Arellano, M. (1996) "Symmetrically normalised instrumental variable estimation using panel data", CEMFI, Working Paper nº 9612, (September).

Alvarez de Toledo, P.; Rojo, J.; Toribio, A. & Usabiaga, C. (2000) "Convergencia: Un análisis conjunto de los sectores. Aplicación al caso de las regiones españolas", FEDEA, Documento de Trabajo 2000-06, (February).

Amemiya, T. (1967) "A note on the estimation of Balestra-Nerlove models", Technical Report nº 4, Institute for Mathematical Studies in Social Sciences, Stanford University.

Anderson, T. W. & Hsiao, C. (1981) "Estimation of dynamic models with error components", *Journal of the American Statistical Association*, 76, 598-606.

Anderson, T. W. & Hsiao, C. (1982) "Formulation and estimation of dynamic models using panel data", *Journal of Econometrics*, 18, 47-82.

Andrés, J. & Doménech, R. (1995) "La convergencia real en Europa", Dirección General de Planificación, Secretaría de Estado de Hacienda, Ministerio de Hacienda, Documento de Trabajo D-95010, (Diciembre).

Andrés, J.; Doménech, R. & Molinas, C. (1996) "Macroeconomic performance and convergence in OECD countries", *European Economic Review*, 40, 9, (December), 1683-1704.

Andrews, D. W. K. (1991) "Heteroscedasticity and autocorrelation consistent covariance matrix estimation", *Econometrica*, 59, 817-858,

Anscombe, F. J. (1967) "Topics in the investigation of least squares (with discussion)", *Journal of the Royal Statistical Society*, Series B, 1-52.

Arellano, M. (1988) "An alternative transformation for fixed effects models with predetermined variables", Applied Economics Discussion Paper nº 57, Institute of Economics and Statistics, University of Oxford.

Arellano, M. (1989) “A note on the Anderson-Hsiao estimator for panel data”, *Economics Letters*, 31, 337-341.

Arellano, M. & Bond, S. (1991) “Some tests of specification for panel data: Monte carlo evidence and an application to employment equations”, *Review of Economic Studies*, 58, 277-297.

Arellano, M, & Bover, O. (1995) “Another look at the instrumental variable estimation of error-components model”, *Journal of Econometrics*, 68, 29-51.

Atkinson, A. B. (1970) “On the measurement of inequality”, *Journal of Economic Theory*, 3, 244-263.

Azariadis, C. & Drazen, A. (1990) “Threshold externalities in economic development”, *Quarterly Journal of Economics*, 109, 2, (May), 465-490.

BBV (varios años) *Renta Nacional de España y su Distribución Provincial*, Banco de Bilbao y Banco Bilbao-Vizcaya.

Balestra, P. (1992a) “Introduction to linear models for panel data”, Cap.- 2 in L. Mátyás & P. Sevestre (Eds.) *The Econometrics of Panel Data*, Kluwer Academic Publishers, 19-29.

Balestra, P. (1992b) “Fixed effect models and fixed coefficient models”, Cap.- 3 in L. Mátyás & P. Sevestre (Eds.) *The Econometrics of Panel Data*, Kluwer Academic Publishers, 30-45.

Balestra, P. & Nerlove, M. (1966) “Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas”, *Econometrica*, 34, 585-612.

Balgati, B. (1995) *Econometric Analysis of Panel Data*, John Wiley & Sons Ltd, New York.

Banerjee, A. (1999) “Panel data unit roots and cointegration: An overview”, *Oxford Bulletin of Economics and Statistics*, 61, Special Issue, (November), 607-629.

Banerjee, A.; Marcellino, M. & Osbat, C. (2000) “Some cautions on the use of panel methods for integrated series of macro-economic data”, European University Institute, EUI Working Papers, ECO nº 2000/20, (November).

Barro, R. J. (1991) “Economic growth in a cross section of countries”, *The Quarterly Journal of Economics*, 106, (May), 407-443.

Barro, R. J. (1999) “Inequality, growth and investment”, NBER Working Paper 7038, (March).

Barro, R. J. & Lee, J. W. (1994a) "Losers and winners in economic growth", *Proceedings of the World Bank Annual Conference on Development Economics*, Washington D. C., World Bank, 267-297.

Barro, R. J. & Lee, J. W. (1994b) "Sources of economic growth", *Carnegie Rochester Conference on Public Policy*, 40, 1-46.

Barro, R. J. & Sala-i-Martín, X. (1991) "Convergence across states and regions", *Brookings Papers on Economic Activity*, 1, (April), 107-182.

Barro, R. J. & Sala-i-Martín, X. (1992) "Convergence", *Journal of Political Economy*, 100, 2, 223-251.

Barro, R. J. & Sala-i-Martín, X. (1995) *Economic Growth*, McGraw Hill, New York.

Baumol, W. J. (1986) "Productivity growth, convergence, and welfare", *American Economic Review*, 76, 5, (December), 1072-1085.

Baumol, W. J.; Blackman, S. A. B. & Wolff, E. N. (1989) *Productivity and American Leadership: The long view*. M.I.T. Press, Cambridge and London.

Benhabib, J. & Spiegel, M. M. (1997) "Cross-country growth regressions", Working Paper 97-20, CV Starr Center, New York University.

Bernard, A. B. (1992) "Empirical implications of the convergence hypothesis", Working Paper, Economics Department, MIT. Cambridge, MA.

Bernard, A. B. & Durlauf, S. N. (1991) "Convergence of international output movements", *National Bureau of Economic Research*, Working Paper 3717, (May).

Bernard, A. B. & Durlauf, S. N. (1995) "Convergence of international output", *Journal of Applied Econometrics*, 10, 97-180.

Bernard, A. B. & Durlauf, S. N. (1996) "Interpreting tests of convergence hypothesis", *Journal of Econometrics*, 71, 1/2, (March/April), 161-173.

Bhargava, A. & Sargan, J. D. (1983) "Estimating dynamic random effects models from panel data covering short time periods", *Econometrica*, 51, 6, 1635-1659.

Binder, M.; Hsiao, C. & Pesaran, M. H. (2000) "Estimation and inference in short panel vector autoregressions with unit roots and cointegration", Mimeo, Department of Applied Economics, Cambridge University, (April).

Binder, M. & Pesaran, M. H. (1996) "Stochastic growth", Department of Economics, Working Paper 96-118, University of Maryland.

Blanchard, O. J. & Fisher, S. (1989) *Lectures on Macroeconomics*, MIT Press, Cambridge, Massachusets.

Blundell, R. & Bond, S. (1998) "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics*, 87, 115-143.

Boscá, J. E. (1996) *Crecimiento económico y convergencia en la OCDE. 1960-1990.* Tesis Doctoral. Universidad de Valencia.

Boumahdi, R. & Thomas, A. (1991) "Testing for unit roots using panel data. Application to the French stock market efficiency", *Economics Letters*, 37, 1, (September), 77-79.

Box, G. E. P. & Cox, D. R. (1964) "An analysis of transformations", *Journal of the Royal Statistical Society*, Series B, 26, 211-243.

Boyle, G. E. & McCarthy, T. G. (1997) "A simple measure of β-convergence", *Oxford Bulletin of Economics and Statistics*, 59, 2, (May), 257-264.

Breitung, J. & Meyer, W. (1994) "Testing for unit roots in panel data: Are wages on different bargaining levels cointegrated?", *Applied Economics*, 26, 353-361.

Campbell, J. Y. (1994) "Inspecting the mechanism. An analytical approach to the stochastic growth model", *Journal of Monetary Economics*, 33, 463-506.

Campbell, J. Y. & Perron, P. (1991) "Pitfalls and opportunities: What macroeconomicst should know about unit roots", with comments by J. H. Cochrane and J. A. Miron, NBER *Macroeconomics Annual*, 141-219.

Canjels, E. & Watson, M. W. (1997) "Estimating deterministic trends in the presence of serially correlated errors", *The Review of Economics and Statistics*, 79, 2, (May), 184-200.

Canova, F. & Marcet, A. (1995) "The poor stay poor: Non-convergence across countries and regions", Discussion Paper 1265, CEPR, (November).

Cantó, O. (2000) "Income mobility in Spain: How much is there?", *Review of Income and Wealth*, 46, 1, (March), 85-101.

Carlino, G. A. & Mills, L. O. (1993) "Are U.S. regional income converging? A time series analysis", *Journal of Monetary Economics*, 32, 335-346.

Carree, M. & Klomp, L. (1997) "Testing the convergence hypothesis: A comment", *The Review of Economics and Statistics*, 79, 683-686.

Caselli, F.; Esquivel, G. & Lefort, F. (1996) "Reopening the convergence debate: A new look at cross-country growth empirics", *Journal of Economic Growth*, 1, (September), 363-389.

Cass, D. (1965) "Optimum growth in an agregative model of capital accumulation", *Review of Economic Studies*, 32, (July), 233-240.

Chakravarty, S. R. (1990) *Ethical Social Index Numbers*, Springer Verlag, Berlin.

Chamberlain, G. (1982) "Multivariate regression models for panel data", *Journal of Econometrics*, 18, 5-46.

Chamberlain, G. (1984) "Panel data", in Z. Grilinches & M. D. Intriligator (Eds.), *Handbook of Econometrics*, Vol.-2, Elsevier Publisher, Amsterdam, 1247-1313.

Chiang, A. C. (1984) *Fundamental Methods of Mathematical Economics*, Third edition, International Student Edition, McGraw-Hill Book Company, London.

Cho, D. (1996) "An alternative interpretation of conditional convergence results", *The Journal of Money, Credit and Banking*, 28, 4, (November), 669-681.

Coakley, J. & Fuertes, A. M. (1997) "New panel unit root tests of PPP", *Economics Letters*, 57, 17-22.

Cosslett, S. R. (1993) "Estimation from endogenously stratified samples", in G.S. Maddala, C. R. Rao & Vinod, H. D. (Eds.) *Handbook of Statistics*, Volume 11, Amsterdam, North-Holland, 1-43.

Costello, D. (1993) "A cross-country, cross-industry comparison of productivity growth", *Journal of Political Economics*, 101, 207-222.

Cowell, F. (1995) *Measuring Inequality*, 2nd Edition, LSE Handbooks in Economics, Prentice Hall, London. (1st. Edition 1977, Phillip Allan Publishers Limited, London).

Culver, S. E. & Papell, D. H. (1997) "Is there a unit root in the inflation rate? Evidence from sequential break and panel data models", *Journal of Applied Econometrics*, 12, 4, 435-444.

Dalton, H. (1920) "The measurement of inequality of income", *Economic Journal*, 30, 348-361.

Davis, S. J.; Haltiwanger, J. & Schuh, S. (1993) "Small business and job creation: Dissecting the myth and reassesing the facts", National Bureau of Economic Research, Working Paper 4492, (October).

Deaton, A. (1987) "Life-cycle models of consumption: Is the evidence consistent with the theory", in T. F. Bewley (Ed.) *Advances in Econometrics*. *Fifth World Congress,* Volume II, Cambridge University Press, Cambridge, 121-148.

Deaton, A. (1997) *The Analysis of Household Surveys. A Microeconometric Approach to Development Policy*. Published for the World Bank. The Johns Hopkins University Press. Baltimore and London.

Deaton, A. & Muellbauer, J. (1980) *Economics and Consumer Behavior*, Cambridge University Press, Cambridge.

De Jong, R. M. & Davidson, J. (2000) "Consistency of kernel estimators of heteroscedastic and autocorrelated convariance matrices", *Econometrica*, 68, 2, (March), 407-423.

De la Fuente, A. (1997) "The empirics of growth and convergence: A selective review", *Journal of Economics Dynamics and Control*, 21, 1, (January), 23-73.

De la Fuente, A. (1998a) "Algunas técnicas para el análisis de la convergencia con una aplicación a las regiones españolas", Dirección General de Análisis y Programación Presupuestaria, D-98007, (Abril).

De la Fuente, A. (1998b) "Whan kind of regional convergence", Dirección General de Análisis y Programación Presupuestaria, D-98010, (Junio).

DeLong, J. B. (1988) "Productivity growth, convergence, and welfare: A comment", *American Economic Review*, 78, 5, (December), 1138-1155.

DeLong, J. B. & Summers, L. H. (1988) "On the existence and interpretation of a 'unit root' in U.S. GNP", National Bureau of Economic Research, Working Paper 2716, (September).

den Haan, W. J. (1995) "Convergence in stochastic growth models. The importance of understanding why income levels differ", *Journal of Monetary Economics*, 35, 65-82.

Dickey, D. A. & Fuller, W. A. (1979) "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American Statistical Association*, 74, 366, (June), 427-431.

Dickson, J. D. H. (1886) "Appendix" to Galton (1886b), *Proceedings of the Royal Society of London*, 40, 63-66.

Diebold, F. X. & Nerlove, M. (1990) "Unit roots in economic time series: A selective survey", in Fomby and Rodes (Eds.) *Advances in Econometrics*, JAI Press Inc., Volume 8, 3-69.

Doppelhofer, G.; Miller, R. & Sala-i-Martín, X. (2000) "Determinants of long-term growth: Robustness tests and model averaging", *Euroconference on innovation, economic growth and european regional cohesion*, Universitat Pompeu Fabra, Barcelona, June 5-6.

Dougherty, C. (1992) *Introduction to Econometrics*, Oxford University Press, Oxford.

Durlauf, S. N. (1993) "Nonergodic economic growth", *Review of Economic Studies*, 60, 2, (April), 349-366.

Durlauf, S. N. (1996) "On the convergence and divergence of growth rates: An introduction", *The Economic Journal*, 106, 437, 1016-1018.

Durlauf, S. N. & Johnson, P. A. (1995) "Multiple regimes and cross-country growth behavior", *Journal of Applied Econometrics*, 10, 4, (October), 365-384.

Durlauf, S. N. & Quah, D. (1998) "The new empirics of economic growth", National Bureau of Economic Research, Working Paper 6422, (February).

DuMouchel, W. H. & Duncan, G. J. (1983) "Using sample survey weights in multiple regression analysis of stratified samples", *Journal of the American Satistical Association*, 78, 535-543.

Easterly, W.; Kremer, M.; Pritchett, L. & Summers, L. H. (1993) "Good policy or good luck?. Country growth performance and temporary shocks", *Journal of Monetary Economics*, 32, 459-483.

Engle, R. F.; Hendry, D. F. & Richard, J. F. (1983) "Exogeneity", *Econometrica*, 51, 2, (March), 277-304.

Entorf, H. (1997) "Random walks with drifts: Nonsense regression and spurious fixed-effect estimation", *Journal of Econometrics*, 80, 2, (October), 287-296.

Esteban, J. M. (1996) "Desigualdad y polarización. Una aplicación a la distribución interprovincial de la renta en España", *Revista de Economía Aplicada*, 4, 11, (Otoño), 5-26.

Esteban, J. M. & Ray, D. (1993) "El concepto de polarización y su medición", en *Igualdad y Distribución de la Renta y la Riqueza*, vol.-2, Fundación Argentaria, Madrid, 1-35.

Esteban, J. M. & Ray, D. (1994) "On the measurement of polarization", *Econometrica*, 62, 819-852.

Evans, P. (1996) "Using cross-country variances to evaluate growth theories", *Journal of Economic Dynamics and Control*, 20, 1027-1049.

Evans, P. (1997) "How fast do economies converge?", *The Review of Economics and Statistics*, 129, 2, (May), 219-225.

Evans, P. & Karras, G. (1996a) "Convergence revisited", *Journal of Monetary Economics*, 37, 249-265.

Evans, P. & Karras, G. (1996b) "Do economies converge? Evidence from a panel of U.S. states", *The Review of Economics and Statistics*, 78, 3, (August), 384-388.

Fingleton, B. (1997) "Specification and testing of Markov chain models: An aplication to convergence in the European union", *Oxford Bulletin of Economics and Statistics*, 59, 3, (August), 385-403.

Fingleton, B. (1999a) "Estimates of time to economic convergence: An analysis of regions of the european union", *International Regional Science Review*, 22, 1, (April), 5-34.

Fingleton, B. (1999b) "Economic geography with spatial econometrics: A 'third way' to analyse economic development and 'equilibrium', with application to the EU regions", *European University Institute, Working Paper* ECO 99/21, (May).

Fingleton, B. (1999c) "Spurious spatial regression: Some monte carlo results with a spatial unit root and spatial cointegration", *Journal of Regional Science*, 39, 1, 1-19.

Fisher, R. A. (1956) *Statistical Methods amd Scientific Inference*, Oliver and Boyd, Edinburgh.

Florax, R. J. G. M. & Rey, S. J. (1995) "The impact of misspecified spatial interaction in linear regression models", in L. Anselin & R. J. G. M. Florax (Eds.) *New Directions in Spatial Econometrics*, Springer-Verlag, Berlin, 111-135.

Forbes, K. J. (1998) "A reassessment of the relationship between inequality and growth, MIT Working Paper, (September).

Foster, J. E. & Ok, E. A. (1999) "Lorenz dominance and the variance of logarithms", *Econometrica*, 67, 4, (July), 901-907

Friedman, M. (1992) "Do old fallacies ever die?", *Journal of Economic Literature*, 30, (December), 2129-2132.

Frisch, R. & Waugh, F. (1933) "Partial time regressions as compared with individual trends", *Econometrica*, 1, 1, (January), 387-401.

Fuller, W. A. (1976) *Introduction to Statistical Time Series*, John Wiley & Sons, New York.

Galton, F. (1869) *Hereditary Genius: An Inquiry Into its Laws and Consequences*, MacMillan, London. (2nd. edition 1892).

Galton, F. (1877) "Typical laws of heredity", *Nature*, 15, 492-495, 512-514, 532-533. También publicado en *Proceedings of the Royal Institution of Great Britian*, 8, 282-301.

Galton, F. (1885) "Section H; Anthropology; Opening address", *Nature*, 32, 507-510.

Galton, F. (1886a) "Regression towards mediocrity in hereditary stature", *Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 246-263.

Galton, F. (1886b) "Family likeness in stature", *Proceedings of the Royal Society of London*, 40, 42-73.

Galton, F. (1888) "Co-relations and their measurement, chiefly from anthropometric data", *Proceedings of the Royal Society of London,* 45, 135-145.

Galton, F. (1889) *Natural Inheritance*, MacMillan, London.

Galton, F. (1908) *Memoirs of My Life*, Methuen, London.

García-Milá, T. & Marimón, R. (1996) "Integración regional e inversión pública en España", en R. Marimón (Ed.) *La Economía Española: Una Visión Differente*, Cap.- 7, Antoni Bosch editor, Barcelona.

Gaulier, G.; Hurlin, C. & Jean-Pierre, P. (1999) "Testing convergence: A panel data approach", *Annales d'Economie et de Statistique*, 55/56, (September/December), 411-427.

Granger, C. W. J. & Hyung, N. (1999) "Spurious stochastics in a short time-series panel data", *Annales d'Economie et de Statistique*, 55/56, (September/December), 299-315.

Green, W. H. & Seaks, T. G. (1991) "The restricted least squares estimator: A pedagogical note", *The Review of Economics and Statistics*, 73, 2, (August), 563-567.

Goerlich, F. J. (1998) "Dinámica de la distribución provincial de la renta. I: Un enfoque desde la óptica de la desigualdad", *Quaderns de Treball* Núm. 69 (nova època), Facultat de Ciències Econòmiques y Empresarials, Universitat de València.

Goerlich, F. J. (2000a) "Desigualdad, diversidad y convergencia: (Más) instrumentos de medida -Estadística descriptiva-", *Monografía*, Instituto Valenciano de Investigaciones Económicas, (Abril). (http://www.ivie.es).

Goerlich, F. J. (2000b) "Dinámica de la distribución provincial de la renta. II: La forma externa de la distribución -Evolución histórica-". Documento de Trabajo DT00-08 a DT00-11, 4 volúmenes. Departamento de Análisis Económico. Universidad de Valencia. (http://www.uv.es/~goerlich).

Goerlich, F. J. (2001a) "Dinámica de la distribución provincial de la renta. III: Movilidad intra-distribucional -Evolución histórica-". Manuscrito en elaboración. Universidad de Valencia.

Goerlich, F. J. (2001b) "Dinámica de la distribución provincial de la renta. IV: Posibles explicaciones y factores condicionantes". Manuscrito en elaboración. Universidad de Valencia.

Goerlich, F. J. & Mas, M. (1998) "Medición de la desigualdad: Variables, indicadores y resultados", *Moneda y Crédito*, 207, (Noviembre), 59-86.

Gould, D. M. & Ruffin, R. J. (1993) "What determines economic growth?", *Economic Review*, Federal Reserve Bank of Dallas, Second Quarter, 25-40.

Griffith, D. (1996) "Some guidelines for specifying the geographic weights matrix contained in spatial statistical models", in S. Arlinghaus, S. & Griffith, D. (Eds.) *Practical Handbook of Spatial Statistics*, Boca Raton, FL, CRC Press, 65-82.

Hall, R. E. & Jones, C. I. (1996) "The productivity of nations", National Bureau of Economic Research, Working Paper 5812, (November).

Hall, R. E. & Jones, C. I. (1997) "Levels of economic activity across countries", *American Economic Review*, Papers and Proceedings, 87, 2, (May), 173-177.

Hall, R. E. & Jones, C. I. (1999) "Why do some countries produce so much more output per worker than others?", *The Quarterly Journal of Economics*, 114, 1, (February), 83-116.

Hall, S.; Lazarova, S. & Urga, G. (1999) "A principal components analysis of common stochastic trends in heterogeneous panel data: Some monte carlo evidence", *Oxford Bulletin of Economics and Statistics*, 61, Special Issue, (November), 749-767.

Hall, S. G.; Robertson, D. & Wickens, M. R. (1992) "Measuring convergence of the EC economies", *The Manchester School*, LX, supplement, (June), 99-111.

Hansen, B. E. (1992) "Consistent covariance matrix estimation for dependent heterogeneous processes", *Econometrica*, 60, 967-972.

Hansen, B. E. (2000) "Sample splitting and threshold estimation", *Econometrica*, 68, 3, (May), 575-603.

Hansen, L. P. (1982) "Large sample properties of generalized method of moments estimators", *Econometrica*, 50, 4, (July), 1029-1054.

Harris, R. D. F. & Tzavalis, E. (1999) "Inference for unit roots in dynamic panels where the time dimension is fixed", *Journal of Econometrics*, 91, 201-226.

Hart, P. E. (1995) "Galtonian regression across countries and the convergence of productivity", *Oxford Bulletin of Economics and Statistics*, 57, 3, (August), 287-293.

Hart, P. E. & Prais, S. J. (1956) "The analysis of business concentration: A statistical approach", *Journal of the Royal Statistical Society*, Series A, 119, 2, 150-181.

Hausman, J. A. (1978) "Especification tests in econometrics", *Econometrica*, 46, 6, (November), 1251-1272.

Hendry, D. F. (1995) *Dynamic Econometrics*, Oxford University Press, Oxford.

Holtz-Eakin, D.; Newey, W. & Rosen, H. S. (1988) "Estimating vector autoregressions with panel data", *Econometrica*, 56, 6, (November), 1371-1395.

Hotelling, H. (1933) "Review of *The triumph of mediocrity in business*, by Horace Secrist", *Journal of tha American Statistical Association*, 28, 184, (December), 463-465.

Hsiao, C. (1986) *Analysis of Panel Data*, Cambridge University Press, Cambridge.

Hsiao, C.; Pesaran, M. H. & Tahmiscioglu, A. K. (1999) "Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods", Mimeo, Department of Applied Economics, University of Cambridge, (September).

Hulten, C. & Srinivasan, S. (1999) "Indian manufacturing industry: Elephant or tiger?. New evidence of the Asian miracle", Mimeo. (October).

Imbens, G. Y. & Lancaster, T. (1996) "Efficient estimation and stratified sampling", *Journal of Econometrics*, 74, 289-318.

INE (varios años) *Anuario Estadístico de España.* Instituto Nacional de Estadística, Madrid.

Im, K. S.; Pesaran, M. H. & Shin, Y. (1997) "Testing for unit roots in heterogeneous panels", First version June-1995, Working Paper 9526, Department of Applied Economics, University of Cambridge, Cambridge.

Islam, N. (1995) "Growth empirics: A panel data approach", *The Quartetly Journal of Economics*, 110, 4, (November), 1127-1170.

Islam, N. (1998) "Growth empirics: A panel data approach - A repply", *The Quartetly Journal of Economics*, 113, 1, (February), 325-329.

Jones, C. I. (1995) "Time series tests of endogenous growth models", *The Quarterly Journal of Economics*, 110, 2, (May), 495-525.

Jones, C. I. (1997a) "On the evolution of the World income distribution", *Journal of Economic Perspectives*, 11, 3, (Summer), 19-36.

Jones, C. I. (1997b) "Convergence revisited", *Journal of Economic Growth*, 2, (June), 131-153.

Jones, C. I. & Manuelli, R. E. (1997a) "Endogenuos growth theory: An introduction", *Journal of Economics Dynamics and Control*, 21, 1, (January), 1-22.

Jones, C. I. & Manuelli, R. E. (1997b) "The sources of growth", *Journal of Economics Dynamics and Control*, 21, 1, (January), 75-114.

Kakwani, N. (1997) "Growth rates of per-capita income and aggregate welfare: An international comparison", *The Review of Economics and Statistics*, 79, 2, (May), 201-211.

Kao, C.; Chiang, H. & Chen, B, (1999) "International R&D spillovers: An application of estimation and inference in panel cointegration", *Oxford Bulletin of Economics and Statistics*, 61, Special Issue, (November), 691-709.

Karlsson, S. & Löthgren, M. (2000) "On the power and interpretation of panel unit root tests", *Economics Letters*, 66, 249-255.

Keane, M. P. & Runkle, D. E. (1992) "On the estimation of panel-data models with serial correlation when instruments are not strictly exogenous", *Journal of Business & Economic Statistics*, 10, 1, (January), 1-29. Con comentarios de P. Schmidt, S.C. Ahn y D. Wyhowski; F. Hayashi; T. MaCurdy y G. Chamberlain y una réplica por parte de M. P. Keane y D. E. Runkle.

Kelly, M. (1992) "On endogenous growth with productivity shocks", *Journal of Monetary Economics*, 30, 47-56.

Kennedy, P. (1986) "Interpreting dummy variables", *Review of Economics and Statistics*, 68, 1, (February), 174-175.

Kim, J. H. (1997) "Relationship between the forward and backward representations of the stationary VAR model", *Econometric Theory*, 13, 6, (December), 889-890. Solution in Kim, J. H. (1998), *Econometric Theory*, 14, 5, (October), 691-693.

King, R.; Plosser, C. Y. & Rebelo, S. T. (1988a) "Production, growth and business cycles. I. The basic neoclasical model", *Journal of Monetary Economics*, 21, 2/3, 195-232.

King, R.; Plosser, C. Y. & Rebelo, S. T. (1988b) "Production, growth and business cycles. II. New directions", *Journal of Monetary Economics*, 21, 2/3, 309-341.

King, R.; Plosser, C. Y.; Stock, J. H. & Watson, M. W. (1991) "Stochastic trends and economic fluctuations", *American Economic Review*, 81, 4, (September), 819-840.

King, R. & Rebelo, S. T. (1993) "Transitional dynamics and economic growth in the neoclasical model", *American Economic Review*, 83, 4, (September), 908-931.

Knight, M.; Loayza, N. & Villanueva, D. (1993) "Testing the neoclasical growth model", *IMF Staff Papers*, 40, 512-541.

Kocherlakota, N. R. & Yi, K.-M. (1995) "Can convergence regressions distinguish between exogenous and endogenous growth models?", *Economics Letters*, 49, 211-215.

Koopmans, T. C. (1965) "On the concept of optimal economic growth", in *The Econometric Approach to Development Planning*, Amsterdam, North Holland.

Lee, K.; Pesaran, M. H. & Smith, R. P. (1995) "Growth and convergence: A multi-country empirical analysis of the Solow growth model", Department of Applied Economics, DAE Working Paper 9531, University of Cambridge.

Lee, K.; Pesaran, M. H. & Smith, R. P. (1997) "Growth and convergence in a multi-country empirical stochastic Solow model", *Journal of Applied Econometrics*, 12, 4, (July), 357-392.

Lee, K.; Pesaran, M. H. & Smith, R. P. (1998) "Growth empirics: A panel data approach - A comment", *The Quartetly Journal of Economics*, 113, 1, (February), 319-323.

Leung, C. K. Y. & Quah, D. T. (1996) "Convergence, endogenous growth, and productivity disturbances", *Journal of Monetary Economics*, 38, 535-547.

Levin, A. & Lin, C. F. (1992) "Unit root tests in panel data: Asymptotic and finite sample properties", Department of Economics, University of California, San Diego, Discussion Paper 92-93 (Revised: 1993).

Levin, A. & Lin, C. F. (1993) "Unit root tests in panel data: New results", Department of Economics, University of California, San Diego, Discussion Paper 93-56.

Levin, R. & Renelt, D. (1992) "A sensitivity analysis of cross-country growth regressions", *American Economic Review*, 82, 4, 942-963.

Lichtenberg, F. R. (1994) "Testing the convergence hypothesis", *The Review of Economics and Statistics*, 76, 576-579.

Loayza, N. (1994) "A test of the international convergence hypothesis using panel data", Policy Resarch Working Paper 1333, The World Bank.

López-Bazo, E.; Vaya, E.; Mora, A. J. & Suriñach, J. (1996) "Regional economic dynamics and convergence in Spain and Europe", European Regional Science Association, 36th European Congress, ETH Zurich, Switzerland, 26-30 August.

MacDonald, R. (1996) "Panel unit root tests and real exchange rates", *Economics Letters*, 50, 7-11.

MacKenzie, D. A. (1981) *Statistics in Britain 1865-1930*, Edinburgh University Press, Edinburgh.

Maddala, G. S. (1977) *Econometrics*, McGraw-Hill, International Book Company, New York.

Maddala, G. S. (1999) "On the use of panel data methods with cross-country data", *Annales d'Economie et de Statistique*, 55/56, (September/December), 429-448.

Maddala, G. S. & Wu, S. (1999) "A comparative study of unit root tests with panel data and a new simple test", *Oxford Bulletin of Economics and Statistics*, 61, Special Issue, (November), 631-652.

Magee, L.; Robb, A. L. & Burbidge, J. B. (1998) "On the use of sampling weights when estimating regression models with survey data", *Journal of Econometrics*, 84, 251-271.

Magnus, J. R. & Neudecker, H. (1988) *Matrix Differential Calculus. With Applications in Statistics and Econometrics*, John Wiley & Sons Ltd, New York.

Mankiw, N. G.; Romer, D. & Weil, D. N. (1992) "A contribution to the empirics of economic growth", *Quarterly Journal of Economics*, 107, 2, (May), 407-437.

Marimón, R. & Zilibotti, F. (1996) "¿Por qué hay menos empleo en España?. Empleo "real" vs. empleo "virtual" en Europa", en R. Marimón (Ed.) *La Economía Española: Una Visión Differente*, Cap.- 2, Antoni Bosch editor, Barcelona.

Mátyás, L. & Sevestre, P. (1992, Eds.) *The Econometrics of Panel Data. Handbook of Theory and Applications*, Kluwer Academic Publishers, Dordrecht.

McCallum, B. T. (1993) "Unit roots in macroeconomic time series: Some critical issues", Economic Quarterly, Federal Reserve Bank of Richmond, 79, 2, (Spring), 13-43.

McCoskey, S. & Kao, C. (1998) "A residual-based test of the null of cointegration in panel data", *Econometric Reviews*, 17, 1, 57-84.

McCoskey, S. & Kao, C. (1999) "Testing the stability of a production function with urbanization as a shift factor", *Oxford Bulletin of Economics and Statistics*, 61, Special Issue, (November), 671-690.

Moon, H. R. & Phillips, P. C. B. (1999) "Maximum likelihood estimation in panels with incidental trends", *Oxford Bulletin of Economics and Statistics*, 61, Special Issue, (November), 711-747.

Moon, H. R. & Phillips, P. C. B. (2000) "Estimation of autoregressive roots near unity using panel data", *Econometric Theory*, 16, 6, (December), 927-997.

Nelson, C. R. & Plosser, C. I. (1982) "Trends and random walks in macroeconomic time series: Some evidence and implications", *Journal of Monetary Economics*, 10, 139-162.

Newey, W. K. & West, K. D. (1987) "A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix", *Econometrica*, 55, 703-708.

Nickell, S. (1981) "Biases in dynamic models with fixed effects", *Econometrica*, 49, 1399-1416.

Ogaki, M. (1993) "Unitroots in macroeconometrics: A survey", Bank of Japan Monetary and Economic Studies, 11, 2, (November), 131-154.

Oh, K. Y. (1996) "Purchasing power parity and unit root tests using panel data" *Journal of International Money and Finance*, 15, 405-418.

Paci, R. (1997) "More similar and less equal: Economic growth in the European regions", *Weltwirtschaftliches Archiv*, 133, 4, 609-634.

Paci, R. & Pigliaru, F. (2000) "Technological cath-up and regional convergence in Europe", Mimeo, (January). Presentado en las *I Jornadas de Economía Fundación Caixa Galicia, Crecimiento y Convergencia Regional*, Santiago de Compostela 16 y 17 de Junio de 2000.

Pan, Z. & LaSage, J. P. (1995) "Using spatial contiguity as prior information in vector autoregressive models", *Economics Letters*, 47, 137-142.

Papell, D. H. (1997) "Serching for stationarity: Purchasing power parity under the current float", *Journal of International Economics*, 43, 313-332.

Pearson, K. (1894) "Contributions to the mathematical theory of evolution I. On the dissection of asymetrical frequency curves", *Philosophical Transactions of the Royal Society of London*, series A, 185, 71-110.

Pearson, K. (1895) "Contributions to the mathematical theory of evolution II. Skew variation in homogeneous material", *Philosophical Transactions of the Royal Society of London*, series A, 186, 343-414.

Pearson, K. (1896) "Contributions to the mathematical theory of evolution III. Regression, heredity and panmixia", *Philosophical Transactions of the Royal Society of London*, series A, 187, 253-318.

Pedroni, P. (1997) "Asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis", Mimeo, Indiana University, (April).

Pedroni, P. (1999a) "Critical values for cointegration tests in heterogeneous panels with multiple regressors", *Oxford Bulletin of Economics and Statistics*, 61, Special Issue, (November), 653-670.

Pedroni, P. (1999b) "Fully modified OLS for heterogeneous cointegrated panels", Mimeo, Indiana University, (December).

Perron, P. (1989) "The great crash, the oil price shock, and the unit root hypothesis", *Econometrica*, 57, 6, (November), 1361-1401.

Perron, P. (1990) "Testing for a unit root in a time series with a changing mean", *Journal of Business and Economic Statistics*, 8, 2, (April), 153-162.

Perron, P. & Vogelsang, T. J. (1992) "Testing for a unit root in a time series with a changing mean: Corrections and extensions", *Journal of Business and Economic Statistics*, 10, 4, (October), 467-470.

Perron, P. & Vogelsang, T. J. (1993) "Erratum" - Perron, P. (1989 - *Econometrica*) "The great crash, the oil price shock, and the unit root hypothesis", *Econometrica*, 61, 1, (January), 248-249.

Pesaran, M. H. & Smith, R. (1995) "Estimating long-run relationships from dynamic heterogeneous panels", *Journal of Econometrics*, 68, 79-113.

Phillips, P. C. B. & Moon, H. R. (1999) "Linear regression limit theory for nonstationary panel data", *Econometrica*, 67, 5, (September), 1057-1111.

Phillips, P. C. B. & Perron, P. (1988) "Testing for a unit root in time series regression", *Biometrika*, 75, 2, (June), 335-346.

Pigou, A. C. (1912) *The Economic of Welfare*, London. (Editado por MacMillan, New York en 1952).

Porter, T. M. (1986) *The Rise of Statistical Thinking 1820-1900*, Princeton University Press, Princeton.

Prais, S. J. (1958) "The statistical conditions for a change in business concentration", *Review of Economics & Statistics*, 40, 268-272.

Pritchett, L. (1997) "Divergence, big time", *Journal of Economic Perspectives*, 11, 3, (Summer), 3-17.

Quah, D. (1987) "What do we learn from unit roots in macroeconomics time series?", National Bureau of Economic Research, Working Paper 2450, (December).

Quah, D. (1990) "International patterns of growth: I. Persistence in cross-country disparities", Mimeo. Economics Department, MIT. Cambridge, MA.

Quah, D. (1993a) "Galton's fallacy and test of the convergence hypothesis", *The Sandinavian Journal of Economics*, 95, 4, (December), 427-443.

Quah, D. (1993b) "Empirical cross-section dynamics in economic growth", *European Economic Review*, 37, 2/3, (April), 426-434.

Quah, D. (1994a) "One business cycle and one trend from (many,) many disaggregates", *European Economic Review*, 38, 605-613.

Quah, D. (1994b) "Exploiting cross-section variation for unit root inference in dynamic data", *Economics Letters*, 44, 9-19.

Quah, D. (1996a) "Twin peaks: Growth and convergence in models of distribution dinamics", *Economic Journal*, 106, 437, (July), 1045-1055.

Quah, D. (1996b) "Ideas determining convergence clubs", Working Paper, Economics Department, LSE. (April).

Quah, D. (1996c) "Regional cohesion from local isolated actions: I. Historical outcomes." Working Paper, Economics Department, LSE. (December).

Quah, D. (1996d) "Regional convergence clusters across Europe", *European Economic Review*, 40, 3/5, (April), 951-958.

Quah, D. (1996e) "Empirics for economic growth and convergence", *European Economic Review*, 40, 1353-1375.

Quah, D. (1997) "Empirics for growth and distribution: Stratification, polarization, and convergence clubs", *Journal of Economic Growth*, 2, (March), 27-59.

Quah, D. & Sargent, T. J. (1993) "A dynamic index model for large cross-sections", in J. Stock & M. Watson (Eds.) *New Research in Business Cycles, Indicators, and Forecasting*. University of Chicago Press, Chicago.

Rabadan, I. & Salas, R. (1996) "Convergencia y redistribución intertemporal en España: Efecto de los impuestos directos, cotizaciones sociales y transferencias", *Economía Pública*, (Septiembre), Fundación BBV.

Raymond, J. L. & García-Greciano, B. (1994) "Las disparidades en el PIBpc entre las CCAA y la hipótesis de convergencia", *Papeles de Economía Española*, 59, 38-58.

Sala-i-Martín, X. (1990) *On growth and states*, Ph.D. dissertation, Harvard University, Cambridge, MA.

Sala-i-Martín, X. (1994) "Cross-sectional regressions and the empirics of economic growth", *European Economic Review*, 38, 739-747.

Sala-i-Martín, X. (1996) "Regional cohesion: Evidence and theories of regional growth and convergence", *European Economic Review*, 40, 1325-1352.

Sargan, J. D. (1958) "The estimation of economic relationships using instrumental variables", *Econometrica*, 26, 393-415.

Sargan, J. D. (1988) "Testing for misspecification after estimating using instrumental variables", in E. Maasoumi (Ed.) *Contributions to Econometrics: John Denis Sargan*, Vol.-1, Cambridge University Press, Cambridge.

Seber, G. A. F. (1977) *Linear Regression Analysis*, John Wiley & Sons, New York.

Secrist, H. (1933) *The Triumph of Mediocrity in Business*, Chicago.

Selden, T. M. (1994) "Weighted generalized least squares estimation for complex survey data", *Economics Letters*, 46, 1-6.

Sen, A. (1973) *On Economic Inequality*, Oxford University Press, Oxford.

Sevestre, P. & Trognon, A. (1985) "A note on autoregressive error components models", *Journal of Econometrics*, 29, 231-245.

Sevestre, P. & Trognon, A. (1992) "Linear dynamic models", Cap.- 6 in L. Mátyás & P. Sevestre (Eds.) *The Econometrics of Panel Data*, Kluwer Academic Publishers, 95-117.

Shorrocks, A. F. (1980) "The class of additively descomposable inequality measures", *Econometrica*, 48, 613-625.

Shorrocks, A. F. (1982) "Inequality decomposition by factor components", *Econometrica*, 50, 193-211.

Shorrocks, A. F. (1984) "Inequality decomposition by population subgroups", *Econometrica*, 52, 1369-1386.

Solow, R. M. (1956) "A contribution to the theory of economic growth", *Quarterly Journal of Economics*, 70, 1, (February), 65-94.

Solow, R. M. (1970) *Growth Theory: An Exposition*, Cambridge Univerity Press, London.

Spanos, A. (1986) *Statistical Foundations of Econometric Modeling*, Cambridge University Press, Cambridge.

Spanos, A. (1999) *Probability Theory and Statistical Inference. Econometric Modeling with Observational Data*, Cambridge University Press, Cambridge.

Stigler, S. M. (1986) *The History of Statistics: The Measurement of Uncertainty Before 1900*, Harvard University Press, Cambridge, MA.

Stockman, A. (1988) "Sectoral and national aggregate disturbances to industrial output in seven european countries", *Journal of Monetary Economics*, 21, 387-409.

Strazicich, M. C.; Co, C. Y. & Lee, J. (2001) "Are shocks to foreign investment in developing countries permanent or temporary? Evidence from panel unit root tests", *Economics Letters*, 70, 3, (March), 405-412.

Suits, D. B. (1984) "Dummy variables: Mechanics vs interpretation", *Review of Economics and Statistics*, 66, 1, (February), 177-180.

Swan, T. W. (1956) "Economic growth and capital accumulation", *Economic Record*, 32, (November), 334-361.

Swamy, P. A. V. B. (1971) *Statistical Inference in Random Coefficient Regression Models*, Lectures Notes in Operations Research and Methematical Systems, 55. Springer-Verlag, Berlin.

Temple, J. (1998) "Robustness tests of the augmented Solow model", *Journal of Applied Econometrics*, 13, 361-375.

Temple, J. (1999) "The new growth evidence", *The Journal of Economic Literature*, 37, 1, (March), 112-156.

Vogelsang, T. J. (1998) "Trend function hypothesis testing in the presence of serial correlation", *Econometrica*, 66, 1, (January), 123-148.

Vogelsang, T. J. & Perron, P. (1998) "Additional tests for a unit root allowing for a break in the trend function at an unknown time", *International Economic Review*, 39, 4, (November), 1073-1100.

White, H. A. (1980) "A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity", *Econometrica*, 48, 4, (May), 721-746.

White, H. A. (1982) "Instrumental variables regression with independent observations", *Econometrica*, 50, 2, (March), 483-499.

Williamson, J. G. (1991) "Productivity and American leadership: A review article", *Journal of Economic Literature*, 29, 51-68.

Wooldridge, J. M. (2001) "Asymptotic properties of weighted M-Estimators for standard stratified samples", *Econometric Theory*, 17, 2, (April), 451-470.

Young, A. (1992) "A tale of two cities: Factor accumulation and technical change in Hong-Kong and Singapore", in *NBER Macroeconomics Annual 1992*, O. J. Blanchard & S. Fisher (Eds.), MIT Press, Cambridge, 13-54.

Young, A. (1995) "The tyranny of numbers: Confronting the statistical realities of the east asian growth experience", *The Quarterly Journal of Economics*, 110, 3, (August), 641-680.

Yule, G. U. (1897) "On the theory of correlation", *Journal of the Royal Statistical Society*, 80, 812-854.

Zietz, J. (2001) "Heteroskedasticity and neglected parameter heterogeneity", *Oxford Bulletin of Economics and Statistics*, 63, 2, (May), 263-273.

Zimmerman, D. J. (1992) "Regression toward mediocrity in economic stature", *American Economic Review*, 82, 3, (June), 409-429.