Dinámica de la Distribución Provincial de la Renta. II:
La Forma Externa de la Distribución –Evolución Histórica-.
(Volumen 2. Sección 6)

Francisco J. Goerlich Gisbert[1]

DT 00-09

Dirección:

Departamento de Análisis Económico
Edificio Departamental Oriental
Universitat de València
Avgda. dels Tarongers, s/n
46022 - VALENCIA
Correo Electrónico: Francisco.J.Goerlich@uv.es
Web: http://www.uv.es/~goerlich
Tel.: 34 – 963828246
Fax: 34 – 963828249

[1] El autor agradece la colaboración de J. C. Robledo y J. M. Sanus en la búsqueda de la información estadística así como los comentarios realizados por M. Mas a una versión de este trabajo, lo que no implica responsabilizarlos de los errores que, a buen seguro, todavía subsisten. Se agradece la financiación recibida de la DGICYT, proyecto SEC98-0895, y del Instituto Valenciano de Investigaciones Económicas (I.V.I.E.).

RESUMEN

Este trabajo es una continuación de Goerlich (1998) y sus aportaciones pueden dividirse en tres grandes bloques. En primer lugar, se ofrece un conjunto amplio de resultados conocidos sobre estadísticos descriptivos cuando estos se calculan a partir de observaciones donde la ponderación de las mismas puede ser una cuestión relevante. En segundo lugar se analiza con cierto detalle, y desde un punto de vista meramente empírico, el concepto de β-convergencia puesto de moda por la moderna teoría del crecimiento económico. Finalmente el trabajo menciona diversos procedimientos para estimar la función de densidad de probabilidad de una variable, centrándose en el método no-paramétrico *kernel*. Se pretende de esta forma ampliar el instrumental analítico ofrecido en Goerlich (1998) para el análisis de la diversidad, la desigualdad y la convergencia. A lo largo del trabajo se ofrece una aplicación práctica con datos provinciales de la renta *per capita* que cubren el periodo 1955-1997. En cierto sentido el trabajo puede considerarse como un panorama muy selectivo de ciertos aspectos de la literatura, convenientemente adaptado a nuestros propósitos.

ABSTRACT

This report is a follow-up of Goerlich (1998) and its contributions can be divided into three big blocks. First, I offer a wide set of known results on descriptive statistics when these statistics are calculated from potentially weighted data and where the weights can be of some relevance. Second, I analyse in detail, and from an empirical point of view, the concept introduced by modern growth theory known as β-convergence. Eventually the report examines different procedures to estimate the probability density function of a variable, focusing on the non-parametric kernel method. In this way I widen the analytical tools offered in Goerlich (1998) for the analysis of diversity, inequality and convergence. An empirical aplication of the different concepts mentioned in the work is offered using *per capita* income data from spanish provinces and the period 1955-1997. In some sense the work can be considered a selective survey of some aspects of the literature.

INDICE

1. **Introdución.**

2. **El periodo de análisis y la variable objeto de estudio.**

3. **Un modelo de referencia:**
   ¿Cuanto explica nuestra ignorancia?

4. **Algunas consideraciones espaciales.**

5. **Estadísticos descriptivos: Convergencia-s.**
   Una primera aproximación a los hechos estilizados.

6. **Convergencia-b y otras historias:**
   Una digresión.

7. **Dinámica de la distribución de la renta *per capita*:**
   Convergencia-d.

8. **A modo de conclusiones.**

Apéndice:

- **Fuentes estadísticas.**
- **Listado de cuadros.**

Referencias.

6 Convergencia-b y otras historias:

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.

La sección anterior ha descrito una serie de estadísticos útiles en la caracterización de la densidad de probabilidad de nuestra variable de interés, $\phi(x)$. Cuando nuestra muestra era dividida en dos **grupos**, provincias ricas *versus* provincias pobres, podíamos examinar si los estadísticos de cada uno de los grupos convengían a lo largo del tiempo en el sentido de la Real Academia Española, es decir si un gráfico de la evolución temporal de dichos estadísticos describía lineas que se dirigían a unirse en un punto; esta era la idea detrás del concepto de m-**convergencia** examinada en el gráfico 12, si los niveles de renta *per capita* media de ambos grupos de provincias se aproximaban o no en el tiempo, evidentemente esta comparación podría realizarse para cada dos provincias, lo que proporcionaría información detallada sobre las provincias convergentes y divergentes, y adicionalmente podríamos efectuar la comparación para otros estadísticos, de hecho el gráfico 13 lo hace para el coeficiente de variación pero en esta caso no resulta claro que lo interesante sea la convergencia entre la dispersión de ambos grupos más que la reducción de esa dispersión dentro de cada uno de los grupos.

Si tomamos nuestra muestra en su conjunto y no realizamos agrupaciones 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 (1985, Cap.-11.1, p.-383), 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** (sección 7), 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))[2], si bien algunos autores insisten en que la dispersión debe disminuir para hablar de convergencia (Lichtenberg (1994), Carree y Klom (1997)).

Este concepto de convergencia, entendido como una reducción en la dispersión *cross-section* a lo largo del tiempo y conocido como **s-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 **b-convergencia**[3]: diremos que existe **b-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 cualquier otra variable) de dichas unidades económicas y su **valor inicial** (Sala-i-Martín (1996), p.-1327), este fenómeno es conocido tambien 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**.

---

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

[3] La terminología **s-convergencia** y **b-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, Roberson y Wickens (1992) o Quah (1993a).

El concepto de b**-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), **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 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 (1997b, 1998b), Jones (1997a,b), Jones y Manuelli (1997a,b), Pritchett (1997), Hall y Jones (1996, 1997, 1999), Durlauf y Quah (1998), Temple (1999)).

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

$$g_{x_i} = \alpha - \beta x_{i,t-1} + u_{it} \tag{29}$$

donde g_{x_i} representa la tasa de crecimiento[4] 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

[4] Normalmente expresada en términos anuales.

de crecimiento del individuo o región *i* en el periodo *t* y que como primera aproximación podemos suponer independiente e idénticamente distribuido, tanto en el corte transversal como en la dimensión temporal[5]. La existencia de **b-convergencia implica b > 0 en (29)**, 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 **b < 1**; como se hará evidente más adelante, **1 < b < 2 implica** una situación en la que se producen **saltos periódicos** (*leapfrogging*) dentro de la distribución *cross-section*, de forma que ricos y pobres alteran su posición relativa periodo a periodo hasta alcanzar el estado estacionario[6], este comportamiento no puede ocurrir en el contexto del modelo neoclásico de crecimiento que dió origen a la ecuación (29) (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 **b > 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 su posición relativa periodo a periodo sin alcanzar ningún estado estacionario[7] (*overshooting*); hay β-convergencia en el sentido definido más arriba, pero no realmente convergencia, una situación de nula relevancia práctica.

Comentario 1: Acerca del cálculo de tasas de crecimiento.

La ecuación (29) 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

---

[5] La cuestión de las ponderaciones, que examinamos con detalle en la sección anterior, tambien es relevante ahora, sin embargo al igual que en las secciones 3 y 4 dichas ponderaciones no serán tenidas en cuenta en este informe. A las complicaciones normales derivadas de tener en cuenta dichas ponderaciones hay que añadir ahora dos más, en primer lugar que la ecuación (29) es normalmente estimada en logaritmos y utilizar las mismas ponderaciones para log *x* que para *x* hace de nuevo perder intuición y sugiere una agregación por medias geométricas en lugar de por medias aritméticas (Attanasio y Weber (1993)); y en segundo lugar que la ecuación (29) es una ecuación dinámica y puesto que las ponderaciones son variables en el tiempo, p_{it}, ello genera discontinuidades en las mismas de difícil solución e interpretación, no obstante una solución aproximada de mínimos cuadrados generalizados siempre sería posible (Green (1993), Sec.-9.4).

[6] En términos de series temporales el proceso estocástico para *x* sería (asintóticamente) **estacionario**, pero presentaría **autocorrelación negativa**.

[7] En términos de series temporales el proceso estacástico para *x* sería **no estacionario**, presentando además **autocorrelación negativa**.

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 tambien el signo de variación.

A modo meramente ilustrativo el cuadro A19 ofrece algunas formas alternativas de calcular tasas de crecimiento entre dos periodos de tiempo, 1955-1997, cuando tenemos información sobre periodos de tiempo intermedios[8]. 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

$$x_t = x_{t-j}(1+g_x)^j \tag{30}$$

de donde se obtiene[9],

$$g_x^1 = \left(\frac{x_t}{x_{t-j}}\right)^{1/j} - 1 = \exp\left\{\frac{1}{j}\left(\log x_t - \log x_{t-j}\right)\right\} - 1 \tag{31}$$

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

$$g_x^2 = \frac{1}{j}\left(\log x_t - \log x_{t-j}\right) \tag{32}$$

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

[8] Consideraremos solamente los años impares ignorando 1994 y 1996.
[9] Este es el cálculo de la función @TASA(•) de Lotus 123™.

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

$$g_x^3 = \frac{\Sigma_{s=1}^{j} g_{x,s}^1}{j} \tag{33}$$

o alternativamente haciendo el cambio de variable $j = t-1$ en (30), $x_t = x_1(1+g_x)^{t-1}$, y tomando logaritmos

$$\log x_t = \log x_1 + (t-1).\log(1+g_x) \tag{34}$$

lo que sugiere la siguiente regresión logarítmica

$$\log x_t = \alpha_0 + \alpha_1 t + \varepsilon_t \tag{35}$$

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 (35) permite, una vez estimado α_1 por $\hat{\alpha}_1$, obtener una estimación de g_x como,

$$g_x^4 = \mathrm{e}^{\hat{\alpha}_1} - 1 \tag{36}$$

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

$$g_x^5 = \hat{\alpha}_1 \tag{37}$$ [10]

La práctica habitual estima (35) 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

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

variación del centro del periodo mientras que las menores ponderaciones se aplican al principio y al final del periodo (Kakwani (1997)), ello que sugiere que otros esquemas de ponderación son posibles y quizá más deseables (Kakwani (1997)). Los resultados del cuadro A19 se han obtenido de esta forma pero obsérvese que g_x tambien aparece en α_0 por lo que podría obtenerse igualmente 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 (35) está probablemente correlacionado.

Los resultados del cuadro A19 muestran que en nuestro caso las diferencias entre los distintos procedimientos de cálculo de las tasas de crecimiento no difieren sustancialmente aunque es posible detectar algunas discrepancias significativas en unas pocas provincias, como por ejemplo Guipúzcoa que cuando utilizamos los métodos de regressión muestran tasas de crecimiento sensiblemente menores a consecuencia del esquema de ponderación implícito en el cálculo de las tasas agregadas y del hecho de que en el periodo 1975-85 esta provincia presentara un crecimiento negativo importante (cuadro A4). Por razones que se harán más evidentes posteriormente en esta sección utilizaremos la aproximación logarítmica g_x^2 como método de cálculo de tasas de crecimiento en la ecuación de 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 (29) queda formulada como

$$\log x_{it} - \log x_{i,t-1} = \alpha - \beta x_{i,t-1} + u_{it} \qquad (38)$$

Comentario 2: Linealidad en la ecuación de b*-convergencia.*

La ecuación que formaliza el concepto de β-convergencia, (29), 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**;

$$g_{x_i} = \alpha - \beta \log x_{i,t-1} + u_{it} \tag{39}$$

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 a continuación; 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, (29) y (39), pueden anidarse en la siguiente formulación general

$$g_{x_i} = \alpha - \beta x_{i,t-1}^{(\lambda)} + u_{it} \tag{40}$$

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

$$x_{i,t-1}^{(\lambda)} = \begin{cases} \dfrac{x_{i,t-1}^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \log x_{i,t-1} & \lambda = 0 \end{cases} \tag{41}$$

Si comparamos para nuestro conjunto de datos y el periodo completo, 1955-1997, la ecuación lineal (29) con la ecuación lineal en logaritmos (39) observaremos que el R^2 para (29) es del 67.38% mientras que para (39) es del 66.96%, lo que favorece ligeramente la especificación lineal, cuando estimamos la ecuación no lineal (40) obtenemos un R^2 ligeramente superior, $R^2 = 67.76\%$, para un valor de λ estimado de 0.59, sin embargo puesto que ahora hemos estimado un parámetro adicional nuestra medida de bondad del ajuste debe tener en cuenta la pérdida de un grado de libertad adicional respecto a las ecuaciones (29) y (39), en este caso si comparamos el coeficiente de determinación corregido, $\overline{R}^2$, de ambas ecuaciones observamos que el de la ecuación no lineal (40) se sitúa a medio camino entre el de las ecuaciones (29) y (39), 66.39% frente a 66.70% y 66.28% respectivamente, por lo dada la magnitud de las diferencias no parece que estas sean significativas; pero lo que es más importante la estimación de λ no es estadísticamente

significativa, lo que permite concluir que $\lambda = 0$, y por tanto la regresión lineal en logaritmos (39) es una aproximación razonable.

Muestra **especificación operativa** para cuantificar el concepto de β-convergencia es por tanto

$$\log x_{it} - \log x_{i,t-1} = \alpha - \beta \log x_{i,t-1} + u_{it} \tag{42}$$

en consonancia con la práctica general. 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.

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

Observamos que (42) puede escribirse como

$$\log x_{it} = \alpha + \rho \log x_{i,t-1} + u_{it} \qquad \rho = 1 - \beta \tag{43}$$

un proceso AR(1) en logaritmos; por lo tanto **b-convergencia, b > 0, implica r < 1** en (43), 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,

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

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 (43)/(42). 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[11].

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)).

[11] No todos los autores estarían sin embargo de acuerdo con esta afirmación (Binder y Pesaran (1996)).

- La **dinámica** en (43), un solo desfase, es probablemente demasiado restrictiva y debe ser ampliada (Evans y Karras (1996a), Evans (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 b-**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 (43) es importante y no puede ser eliminado sin perder una parte sustancial del análisis.

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 (43) donde x_i **es reinterpretada como la renta relativa de dos países o regiones y a = 0** (los paises o regiones comparten el mismo estado estacionario)[12], 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

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

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 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))[13]

$$\lim_{T\to\infty} E\left(\log x_{i,t+T} - \log x_{j,t+T} \middle| \boldsymbol{I}_t\right) = 0 \tag{44}$$

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

$$\lim_{T\to\infty} E\left(\log x_{i,t+T} - \log x_{j,t+T} \middle| \boldsymbol{I}_t\right) = 0 \quad \Rightarrow \quad E\left(\log x_{i,t+T} - \log x_{j,t+T} \middle| \boldsymbol{I}_t\right) < \log x_{it} - \log x_{jt} \tag{45}$$

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[14]. 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 ambas definiciones son útiles ya que representan diferentes implicaciones del modelo, β-convergencia para un T fijo, f-convergencia para $T\to\infty$.

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

Obsérvese que si en (43) 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 (43), 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*-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**, $T \to \infty$, 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 conjunto 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)). 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 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 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 (43) 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)).

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))[15].

(*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 conjuntos de datos algunos autores han tratado de contrastar el concepto de **b-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 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)[16].

---

[15] Raíces unidad en σ-convergencia.

[16] 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 (1995), Entorf (1997), Benerjee (1999), Hall, Lazarova y Urga (1999), Maddala y Wu (1999), McCoskey y Kao (1998), Moon y Phillips (1999), Pedroni (1999a, b), Harris y Tzavalis (1999)) 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), Papell (1997), Coakley y Fuertes (1997), Pedroni (1997), Kao, Chiang y Chen (1999), McCoskey y Kao (1999)).

Muchos autores considerarían la hipótesis H_0: $\rho = 1$ ($\beta = 0$) en (43)/(42) 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), De Long y Summers (1988)).

En nuestro caso la dimensión temporal es excesivamente reducida para llevar a cabo contrastes de raíces unidad para las rentas relativas provinciales dos a dos, ya que aunque nuestra dimensión temporal cubre un periodo de 43 años sólo disponemos de 22 observaciones temporales, $T = 22$, lo que no parece suficiente para realizar inferencia de largo plazo, por muy interesante que el ejercicio fuera, sin embargo si podemos aprovechar la estructura de panel de los datos, ya que $n.T = 1100$ proporciona un número razonablemente elevado de observaciones para poder realizar con cierto margen de confianza un ejercicio de estas características.

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)) existe 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 componentes deterministas (Campbell y Perron (1991)) y en la posible heterogeneidad entre las observaciones *cross-section* (Im, Pesaran y Shin (1995), Maddala y Wu (1999)).

Dado el carácter ilustrativo de nuestros cálculos utilizaremos una aproximación relativamente sencilla al problema de contrastar H_0: $\rho = 1$ en (43) y que deriva de Levin y Lin (1992) como una extensión directa de los contrastes de Dickey y Fuller (1979). La ecuación (43) puede ser rescrita convenientemente como

$$\Delta \log x_{it} = \alpha + (\rho - 1) \log x_{i,t-1} + u_{it} \tag{46}$$

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

Para dotar a (46) 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 (46) de **efectos fijos temporales** es suficientemente flexible para nuestros propósitos[17], al igual que hicimos en el análisis de varianza, (1). En segundo lugar deberemos **relajar la dinámica**, que en la práctica debe ser determinada de forma empírica (Evans y Karras (1996a), Evans (1997)). Finalmente parece restrictivo suponer, al menos inicialmente, que todas las economías tienen el mismo estado estacionario y en consecuencia introduciremos en (46) **efectos fijos individuales** (provinciales). 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

[17] 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)), de esta forma habríamos eliminado las tendencias presentes en los datos y nos centraríamos en el comportamiento de las desviaciones respecto al 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 por tanto no recogen la heterogeneidad en los procesos de crecimiento.

forma de recoger 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 (46) de la siguiente forma

$$\Delta \log x_{it} = \alpha + \lambda_i + \eta_t + (\rho - 1)\log x_{i,t-1} + \Sigma_{j=1}^{p}\theta_j \Delta \log x_{i,t-j} + u_{it} \qquad (47)$$

donde λ_i y η_t tienen la misma interpretación que en (1) 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.

Es importante observar que el contraste de H_0: $\rho = 1$ que vamos a examinar requiere que el coeficiente autoregresivo en (47), ρ, sea homogéneo tanto bajo la hipótesis nula y como bajo la hipótesis alternativa, puesto que agrupa los datos en la dimensión *cross-section*. Im, Pesaran y Shin (1995) 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*. Estos contrastes no serán considerados en nuestro ejemplo, donde disponemos tan solo de 22 observaciones temporales para cada provincia, pero es necesario recordar 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).

Llamemos $\hat{\rho} - 1$ al estimador de mínimos cuadrados ordinarios de $\rho - 1$ en (47) y $t_{\rho=1}$ a 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: $\rho = 1$[18]. Estos autores demuestran que $t_{\rho=1}$ converge asintóticamente en distribución a una normal conforme $n, T \to \infty$ con tal que $\frac{\sqrt{n}}{T} \to 0$, por tanto se permite que la

[18] 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.

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 (47) sean empíricamente blancos, así como de si los efectos fijos temporales son incluidos o no en el modelo y la forma que estos toman[19]. Harris y Tzavalis (1999) derivan las distribuciones de estos estadísticos cuando $n \to \infty$ 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 provincia, log x_{i1}, log x_{i2}, ... log x_{iT}, son relevantes en la estimación del efecto fijo de dicha provincia[20], λ_i; incluir dichos efectos fijos en (47) con $n > 1$ es enteramente análogo a incluir un término constante en (47) 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 (47) podemos utilizar la *ratio F* habitual y considerarla como 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 (47)[21], alternativamente podríamos utilizar una versión del contraste robusta frente a heterocedasticidad de forma desconocida (White

[19] 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)).
[20] Junto con la constante del modelo, α, dado la forma en que los efectos fijos están definidos.
[21] 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 $T \to \infty$ 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.

(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 la distribución en muestras finitas puede diferir apreciablemente de la distribución asintótica (Harris y Tzavalis (1999)) por lo que los niveles de significación en nuestra aplicación práctica se obtuvieron por simulación de Monte Carlo con 10.000 réplicas y a partir de la estimación de los modelos correspondientes bajo H_0 (Evans y Karras (1996a, b)).

Un análisis previo mostró que $p = 2$ en (47) recogía adecuadamente la dinámica de los datos, siendo el segundo desfase significativo a valores superiores al 6.51% utilizando inferencia estándar, por lo que se decidió mantenerlo. Dado este valor de p la estimación de (47) por mínimos cuadrados ordinarios para el periodo 1961 - 1997 produjo los siguientes resultados[22],

$$\Delta \log x_{it} = \hat{\alpha} + \hat{\lambda}_i + \hat{\mu}_t + \underset{(.012374)}{(.910214 - 1)} \log x_{i,t-1} + \Sigma_{j=1}^{p} \hat{\theta}_j \Delta \log x_{i,t-j} + \hat{u}_{it} \qquad R^2 = 0.7478 \tag{48}$$

donde el valor entre paréntesis es el error estándar de $\hat{\rho}$ de mínimos cuadrados ordinarios[23].

La estimación (48) implica un valor de $t_{\rho=1}$ de -7.2558 y cuyo nivel de significación, obtenido por simulación de Monte Carlo[24], resultó ser del 1.15%. Si

[22] Los años 1955, 1957 y 1959 se utilizaron como condiciones iniciales.

[23] Dado que el lapso temporal utilizado en (48) es bianual el valor equivalente de $\hat{\rho}$ para datos anuales es 0.95.

[24] El procedimiento utilizado para determinar empíricamente los niveles marginales de significación de los estadísticos utilizados fue el siguiente:

(*i*) En primer lugar, estimamos por mínimos cuadrados ordinarios el modelo bajo la hipótesis nula que estamos interesados en contrastar, raíz unidad o ausencia de efectos fijos provinciales condicionada en la hipótesis de estacionariedad.

(*ii*) Dado este modelo estimado, generamos 10000 muestras a partir de una distribución normal para u_{it}, suponiendo independencia y homogeneidad. Para ello utilizamos las observaciones de los años 1955, 1957 y 1959 como condiciones iniciales fijas.

utilizamos en (48) errores estándar consistentes frente a heterocedasticidad de forma desconocidad (White (1980) la *t-ratio* obtenida toma un valor de −6.8877, con un nivel de significación empírico en este caso de 1.44%. Así pues la evidencia parece razonablemente robusta, la **renta *per capita* provincial es estacionaria**, al menos una vez hemos tenido en cuenta la posible existencia de efectos fijos provinciales y temporales que recogen de forma flexible la tendencia común a todas las provincias. Los resultados empíricos se muestran pues favorables a la hipótesis de **convergencia**. Este hecho merece probablemente más atención en vista de la aparente omnipresencia de raíces unidad en series temporales macroeconómicas agregadas (Goerlich (1991a, b, 1992)), o al menos una exploración de su robustez mediante la utilización de otros contrastes alternativos de raíces unidad en un contexto de datos de panel.

Finalmente puede resultar de interés analizar si la convergencia en renta *per capita* de las provincias españolas lo es a un solo estado estacionario, **convergencia absoluta o incondicional**, o a múltiples estados estacionarios, **convergencia condicional**. Por lo tanto, condicionado en $\rho < 1$, podemos contrastar la hipótesis H_0: $\lambda_1 = \lambda_2 = \lambda_3 = ... = \lambda_n$ en (48), frente a una alternativa no especificada. El estadístico F-estándar toma un valor F(49,879) = 2.0244, con nivel de significación, tomando como referencia la distribución F-Snedecor, inferior al 0.0001; sin embargo el nivel de significación empírico obtenido mediante la simulación de Monte Carlo fue sensiblemente superior, el 19.26%, mucho más favorable a H_0. Este conflicto entre la distribución asintótica y la de muestras finitas no se produce, sin embargo, si utilizamos para el contraste de H_0 el estadístico χ^2 consistente frente a heterocedasticidad de forma desconocida (White (1980)) que para la ecuación (48)

(*iii*) Para cada una de las muestras generadas estimamos el correspondiente modelo bajo la hipótesis alternativa y calculamos el estadístico deseado, la *t-ratio* o la *F-ratio*. Este procedimiento generó una muestra de 10000 estadísticos de interés.

(*iv*) Los estadísticos obtenidos fueron ordenados de forma ascendente las *t-ratio* y de forma descendente las *F-ratio*.

(*v*) El soporte de las distribuciones fue completado añadiendo un valor −∞ para las *t-ratio* y un valor +∞ para las *F-ratio* como observación 0 y un valor +∞ para las *t-ratio* y un valor 0 para las *F-ratio* como observación 10001.

(*vi*) Finalmente el nivel de significación empírico de cada estadístico se obtuvo como *s*/10002, donde *s* es el entero tal que el estadístico correspondiente se sitúa entre los valores ordenados *s*-ésimo y (*s*+1)-ésimo de los estadísticos simulados.

Para el caso de la inferencia robusta frente a heterocedasticidad de forma desconocida el procedimiento es idéntico excepto que la varianza para la simulación de u_{it} no es constante. En este caso se toman como varianzas una para cada observación de acuerdo con $\hat{u}_{it}^2$, siendo $\hat{u}_{it}$ los resíduos de mínimos cuadrados ordinarios obtenidos de la estimación inical bajo la hipótesis nula.

toma un valor de $\chi^2(49) = 127.2209$, con nivel de significación, tomando como referencia la distribución χ^2, prácticamente nulo, mientras que el nivel de significación empírico obtenido de la simulación de Monte Carlo fue del 0.31%, lo que no deja dudas, en este caso, acerca de la existencia de efectos fijos provinciales. Así pues la evidencia estadística no es totalmente concluyente aunque se muestra más favorable al rechazo de H_0 y en consecuencia los efectos fijos provinciales no desaparecen y concluyendo que la **convergencia** no parece ser absoluta, sino más bien **condicional**. Un tema sobre el que volveremos cuando examinemos con más detalle la estimación de ecuaciones de convergencia aprovechando la estructura de panel de nuestro conjunto de datos.

En **resumen**, nuestro pequeño ejercicio sobre raíces unidad en la renta *per capita* provincial muestra que las provincias españolas si parecen convergen en renta *per capita*, pero que, condicionado en que esta convergencia tiene lugar, la convergencia absoluta puede ser rechazada y lo que observamos es más bien **convergencia condicional**. No obstante la inferencia que podemos deducir de estos resultados estadísticos en términos de la validez del modelo de crecimiento neoclásico, que a continuación analizamos, es de dudosa utilidad ya que como han señalado algunos autores la interpretación en términos de modelos teóricos de resultados estadísticos no es directa (Korchelakota y Li (1995), Binder y Pesaran (1996), Durlauf y Quah (1998, Section 5)).

Comentario 3: Acerca del modelo neoclásico de crecimiento.

La ecuación (42) no es la usualmente utilizada en trabajo aplicado relacionado con el concepto de β-convergencia. En la práctica la **ecuación estimada** es de la forma[25]

$$\frac{1}{j}\left(\log x_{it} - \log x_{i,t-j}\right) = a - \left(\frac{1-e^{-bj}}{j}\right)\log x_{i,t-j} + \varepsilon_{it} \qquad (49)$$

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-

[25] Esta es la ecuación (15) en Barro y Sala-i-Martín (1992, p.-229).

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

El coeficiente sobre $\log x_{i,t-j}$ en la ecuación (49) 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 (49). 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 (49). 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[26].

[26] 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 (49), 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, $\log x_{it} - \log x^* = 0.5\text{x}(\log x_{i,t-j} - \log x^*)$, se obtiene resolviendo por j la ecuación anterior una vez hemos sustituido la distancia que queremos recorrer. En nuestro caso

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

Para futuras referencias resulta útil examinar como la ecuación (46) 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*[27], depende de la log-desviación respecto al estado estacionario. Como es usual esta dependencia es parametrizada por el autovalor negativo de la matríz de primeras derivadas, la negatividad del autovalor asegura la estabilidad del modelo, y 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*, y es equivalente a la condición $\beta < 1$.

Una **cuestión importante** de la argumentación anterior es que el parámetro ***a* en (46) 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 $a = a(w_{it})$ siendo $a(\bullet)$ una función que engloba todas aquellas variables que afectan al estado estacionario de las diferentes economías, w_{it}. Esta puntualización es la que sustenta la diferencia que aparece en la literatura entre **b-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 **b-convergencia**

[27] Ambos conceptos son sin embargo muy diferentes en la vida real, tal y como han señalado Paci (1997) o Goerlich y Mas (1998a).

condicionada, cuando incluimos en la ecuación (49), o alternativamente en la (42), 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 (46) efectos fijos individuales, cuya incorparación al análisis es normalmente recomendable para recoger cualquier tipo de heterogeneidad inobservable.

Como mencionamos en la introducción el análisis de variables condicionantes y su influencia en el proceso de convergencia será tratado con posterioridad[28] (Goerlich (2000b)) por lo que en esta sección **consideraremos sólo** el concepto de b**-convergencia absoluta o incondicional**, y por tanto 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 provincias 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 (42) y (49) es fácil de derivar, a partir de (43) y sustituyendo recursivamente *j* periodos hacia atrás

$$\log x_{it} = \alpha \Sigma_{k=0}^{j-1} \rho^k + \rho^j \log x_{i,t-j} + \Sigma_{k=0}^{j-1} \rho^k u_{i,t-k} \tag{50}$$

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

$$\frac{1}{j}\left(\log x_{it} - \log x_{i,t-j}\right) = \frac{1}{j}\alpha \Sigma_{k=0}^{j-1} \rho^k - \left(\frac{1-\rho^j}{j}\right)\log x_{i,t-j} + \frac{1}{j}\Sigma_{k=0}^{j-1} \rho^k u_{i,t-k} \tag{51}$$

a partir de lo cual observamos que

[28] 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-Martin (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 (42) y el parámetro b en (49), 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 $\mathrm{b} = \mathbf{0}$ $(\rho = 1)$ implica **nula velocidad de convergencia**, es decir ausencia de la misma, por el contrario $\mathrm{b} = \mathbf{1}$ $(\rho = 0)$ implica un ajuste instantáneo, la **convergencia es inmediata** y $\log x_{it}$ carece de correlación serial.

Puesto que (49) implica una estimación no lineal, en ocasiones se estima simplemente esta ecuación en su forma lineal[29]

[29] En ocasiones se estima

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

$$\frac{1}{j}\left(\log x_{it} - \log x_{i,t-j}\right) = a - b' \log x_{i,t-j} + \varepsilon_{it} \tag{52}$$

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 (43) subyacente en (52) 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 (43) 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 (52) 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, p.-1359) es probablemente mejor estimar (49) por métodos no lineales[30] que (52) por métodos lineales[31], de forma que situaciones de hiperconvergencia se muestren como valores extremadamente elevados de *b*.

La razón de la indefinición mencionada en el párrafo anterior radica en que el término e^{-bj} en (49) procede de una log-linearización en **tiempo continuo** y en este caso la

[30] Tal como **mínimos cuadrado no lineales**.
[31] Por ejemplo **mínimos cuadrados ordinarios**.

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 (42) o ρ en (43) es una aproximación de la dinámica de transición del modelo en **tiempo discreto** y en este caso la estabilidad de la relación depende de que $0 < \beta < 2$ o $-1 < \rho < 1$; 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 (43) es estable y no oscilatoria, esto es $0 < \beta < 1$ o $0 < \rho < 1$.

Un **comentario final** que conviene tener presente en lo que hace referencia a la **interpretación teórica de la ecuación de convergencia** (49). 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 contario a estimado mayoritariamente la ecuación (49) 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 $a = a(w_{it})$ en (49) siendo $a(\bullet)$ una función que engloba todas aquellas variables que afectan al estado estacionario, w_{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, y la dimensión *cross-section* del propio concepto de** b**-convergencia** y que no es sino una forma alternativa de ilustrar las aproximaciones *cross-section* y de series temporales 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 es experiencias de crecimiento individuales (Young (1992, 1995)).

Comentario 4: 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 (49) éste parámetro recoge una tendencia temporal derivada de la existencia de progreso técnico (Barro y Sala-i-Martín (1992), p.-230)[32], 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. Como 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 (49) e introduciendo una **tendencia lineal determinista** en *a*, si añadimos incertidumbre **dicha tendencia es fruto de que el progreso 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 (49) colapsa**, **en el sentido de no ser estable** (King, Plosser y Rebelo

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

(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 negative 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 (38)/(49) 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))[33].

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 (38)/(49) y el modelo de crecimiento neoclásico, ni si dicha ecuación es apropiada para discriminar entre modelos (Kocherlakota y Li (1995)). No parece pues

[33] 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.

adecuado utilizar **(49)** 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[34]. Al fin y al cabo 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 no parece apoyar esta hipótesis. 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.

Algunos resultados

La estimación de la ecuación (52) para el conjunto del periodo 1955-1997 y la totalidad de las provincias generó los siguientes resultados,

$$\frac{1}{42}\left(\log x_{i,1997} - \log x_{i,1955}\right) = \underset{(.006445)}{.104991} - \underset{(.001163)}{.012791}\log x_{i,1955} + \hat{\varepsilon}_{i,1997} \qquad R^2 = 0.6696 \tag{53}$$

donde entre paréntesis figuran los errores estándar consistentes frente a heterocedasticidad de forma desconocida (White (1980)) y el R^2 muestra un grado de ajuste aceptable, 66.96%.

[34] La tecnología para aplicar estas técnicas en contextos de datos de panel está todavía en su infancia (Phillips y Moon (1999)).

La **velocidad de convergencia** anual implícita en la estimación de $\hat{b}' = 0.0128$ en (53) es del $\hat{b} = 1.83\%$ con un error estándar robusto frente a heterocedasticidad del 0.25% y por tanto **altamente significativa**[35] y el valor de ρ implícito en esta estimación es de $\hat{\rho} = 0.98$, lo que indica una elevada persistencia fruto de la lenta convergencia hacia el estado estacionario. La heterocedasticidad no parece ser un problema importante[36] de forma que los errores estándares calculados de la forma habitual generaron, en prácticamente todas las estimaciones presentadas en este epígrafe, los mismos resultados cualitativos en términos de inferencia, no obstante se decidió presentar la inferencia robusta frente a la heterocedasticidad por considerarla más conveniente con datos *cross-section* (White (1980)). De esta forma podemos concluir que las **provincias españolas**, para el conjunto del periodo, mostraron **signos de b-convergencia**, es decir **por término medio aquellas provincias que presentaban al principio del periodo una renta *per capita* más elevada**, las provincias ricas de la sección anterior, **han presentado en general menores tasas de crecimiento que la provincias** pobres, aquellas **que presentaban una renta *per capita* menor en 1955**. El gráfico 19 ilustra de forma visual esta situación.

Gráfico 19

El cuadro 16 ofrece la misma estimación para el periodo 1955-1995 y su división en los periodos decenales de referencia, junto con la estimación de b' y su correspondiente error estándar se incluye tambien la información relativa al coeficiente de determinación,

---

[35] Como ya observamos anteriormente un contraste de la hipótesis H_0: $b = 0$ en (49), es decir de que la **velocidad de convergencia es estadísticamente significativa**, lo que es equivalente a H_0: $b' = 0$ en (52) o a H_0: $\beta = 0$ en (42), es tambien equivalente a H_0: $\rho = 1$ en (43), es decir un **contraste de raíz unidad**. En un contexto *cross-section* esto es irrelevante puesto que la dimensión temporal es inexistente y podemos realizar la inferencia por los procedimientos habituales, sin embargo el tema se vuelve relevante en un contexto de datos de panel, aún cuando *T* sea fijo y relativamente pequeño (Harris y Tzavalis (1999)).

Debe observarse, además, que puesto que la relación entre los parámetros b', b, β y ρ es no lineal los contrastes de significatividad anteriores no son numéricamente equivalentes (Gregory y Veall (1985)).

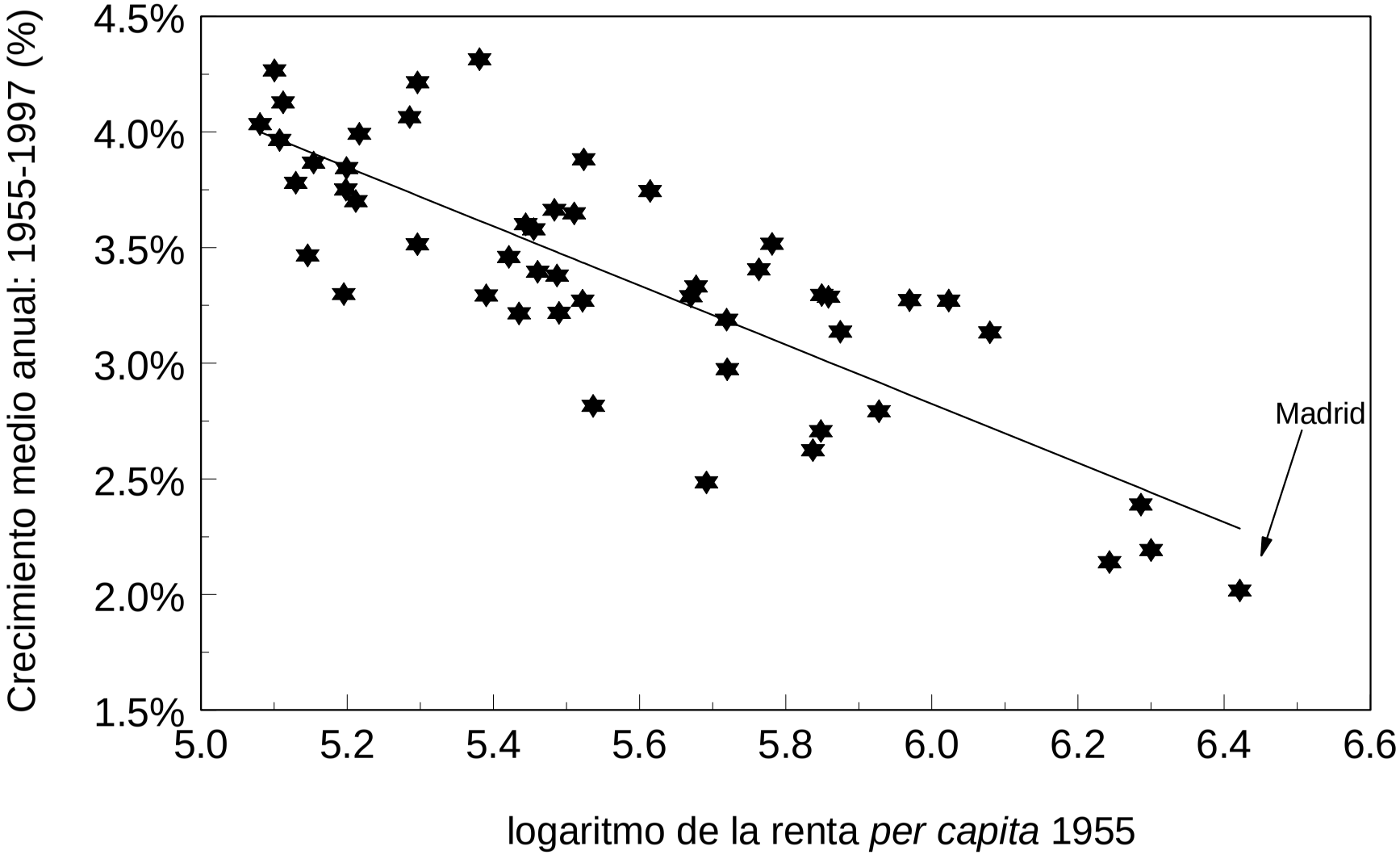
[36] La **heterocedasticidad** fue contrastada en todos los casos mediante un estadístico *F* de la hipótesis H_0: $\gamma = 0$ frente a H_1: $\gamma \neq 0$ en la regresión auxiliar

$$\hat{\varepsilon}_i^2 = \alpha + \gamma \hat{y}_i^2 + u_i$$

donde $\hat{\varepsilon}_i$ son los resíduos de mínimos cuadrados ordinarios de una regresión anterior y $\hat{y}_i$ son los valores ajustados de la variable dependiente en dicha regresión. Dicho contraste es ofrecido por MICROFIT Versión 3.0 (Pesaran y Pesaran (1995)) y puede ser considerado como una versión resumida del contraste de White (1980).

Gráfico 19. *Beta*-convergencia

Periodo: 1955-1997



R^2, y debajo de este el nivel de significación de la hipótesis $H_0{:}b' = 0\ vs\ H_1{:}b' \neq 0$, así como los valores de ρ y la **velocidad de convergencia**, *b*, en términos porcentuales. Podemos observar como el valor estimado de b' es consistentemente negativo y significativo al 5%, sin embargo en el periodo 1985-1995 dicho parámetro no es significativo a niveles inferiores al 1.47%, esta es la única década para la que parece haberse detenido o al menos relentizado el fenómeno de la β-convergencia. En términos de **la velocidad de convergencia**, esta **no parece haber sido estable a lo largo del periodo**, mostrándose mucho más intensa en las dos décadas centrales que al principio y al final del periodo, no obstante **por término medio para la totalidad del periodo una velocidad de convergencia entorno al 2%**, enfatizado por Sala-i-Martín (1996), **no parece una mala aproximación**, a pesar de que este valor no parece que se haya mantenido constante a lo largo del tiempo.

Cuadro 16

La bondad del ajuste, medida en términos del R^2, tambien es notablemente diferente entre periodos, pero destaca el hecho de que el ajuste es mucho mejor para periodos largos, lo que confirma la idea de que la denominada **b-convergencia es un fenómeno de largo plazo** y que en el corto plazo *shocks* transitorios pueden alterar la tendencia de las regiones pobres a crecer por encima de las ricas.

El valor implícito de ρ en las estimaciones del cuadro 16 se sitúa en un entorno entre 0.98-0.99 para el conjunto de provincias, lo que muestra una elevada persistencia en $\log x_{it}$. Este es un valor mucho más estable que la estimación de la velocidad de convergencia, *b*. La razón se debe a que la relación entre ambos parámetros es no-lineal y el intervalo (0, 1) en el que varía ρ es transformado en el intervalo (0, ∞) dentro del que varía *b*; en consecuencia muy pequeñas alteraciones en el valor de ρ son compatibles con velocidades de convergencia aparentemente bastante dispares. En concreto **con una precisión de dos decimales valores de r de 0.98**[37] **son compatibles con velocidades de convergencia situadas entre el 1.61% y el 2.53%**.

[37] Lo que con una precisión de tres decimales equivale en la práctica a valores de ρ en el intervalo [0.975, 0.984].

Cuadro 16: beta-convergencia

Periodo	Provincias: Todas (50)				Provincias: Ricas (15)				Provincias: Pobres (35)			
	b'	**R2**	r	***b***	***b'***	**R2**	r	***b***	***b'***	**R2**	r	***b***
1955-1965	-0.0105	0.2228	0.99	1.11%	-0.0234	0.4140	0.97	2.67%	-0.0049	0.0193	0.99	0.51%
	0.0026	0.0001			0.0052	0.0000			0.0060	0.4093		
1965-1975	-0.0160	0.2922	0.98	1.74%	-0.0239	0.2817	0.97	2.73%	-0.0108	0.0622	0.99	1.15%
	0.0033	0.0000			0.0085	0.0050			0.0051	0.0341		
1975-1985	-0.0212	0.3899	0.98	2.39%	-0.0367	0.3080	0.96	4.57%	-0.0218	0.2679	0.98	2.45%
	0.0039	0.0000			0.0136	0.0071			0.0049	0.0000		
1985-1995	-0.0073	0.0888	0.99	0.76%	-0.0060	0.0188	0.99	0.62%	-0.0064	0.0373	0.99	0.66%
	0.0030	0.0147			0.0116	0.6036			0.0058	0.2698		
1955-1995	**-0.0125**	**0.6147**	**0.98**	**1.73%**	**-0.0206**	**0.6712**	**0.96**	**4.32%**	**-0.0127**	**0.3470**	**0.98**	**1.77%**
	0.0013	**0.0000**			**0.0026**	**0.0000**			**0.0027**	**0.0000**		

Nota: Los errores estándar son consistentes frente a heterocedasticidad de forma desconocida. White (1980).

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 a un simple **sesgo estadístico** debido a la presencia de raíces unidad, e indique en la práctica ausencia de convergencia[38]. 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 (1996) 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[39], pero sesgado a la baja para un *T* finito (Fuller (1976, Cap.-8)), lo que encaja perfectamente con el 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 insesgada de un valor poblacional igual a la unidad.

De hecho si calculamos todas las regresiones *cross-section* posibles de β-convergencia para nuestro conjunto de datos, lo que implica calcular un total de 231 regresiones encontramos que en 150 de las mismas, lo que representa un 65% del total, el valor estimado de ρ es igual a 0.98; tal regularidad merece probablemente ser explorada. Así pues aunque esta es una explicación razonable y que merece más atención por parte de la literatura, las simulaciones de Monte Carlo presentadas por Quah (1996, Sec.-2) a partir de paseos aleatorios (*random walk*) independendientes en el corte transversal muestran que esta no puede ser toda la explicación a la regularidad observada, ya que las simulaciones de Quah (1996, Table 1, p.-1359) muestran una mayor diversidad que la observada en la práctica.

Antes de acabar con estos primeros resultados merece la pena realizar algunos ejercicios adicionales. En primer lugar, **si eliminamos Madrid** de la ecuación de

---

[38] Otros autores han argumentado que este **sesgo estadístico** puede estar causado por la **transformación logarítmica-nolineal** usualmente utilizada en las ecuaciones de convergencia (Quah (1996), nota 3, p.-1356).
[39] En realidad "super-consistente".

convergencia (49), lo que está motivado por el hecho de que esta observación fue detectada como un *outlier* en la sección anterior, observamos que **no cambia prácticamente nada**, la velocidad de convergencia pasa del 1.83% al 1.74%; así pues a pesar de que la contribución de Madrid a la dispersión de la distribución de la renta *per capita* era muy importante al principio del periodo (gráfico 11) no parece que esta única observación sea relevante en términos de β-convergencia.

Por último al igual que hicimos anteriormente dividimos nuestra muestra en dos grupos, las **provincias ricas**, aquellas que en 1955 tenían una renta *per capita* superior a la media nacional, y las **provincias pobres**, aquellas que en 1955 tenían una renta *per capita* inferior a la media nacional, y estimamos la ecuación de convergencia (52) para ambos grupos separadamente, para el total del periodo y por sub-periodos. Los resultados de dicho ejercicio se muestran numéricamente en el cuadro 16[40] y de forma visual en el gráfico 20.

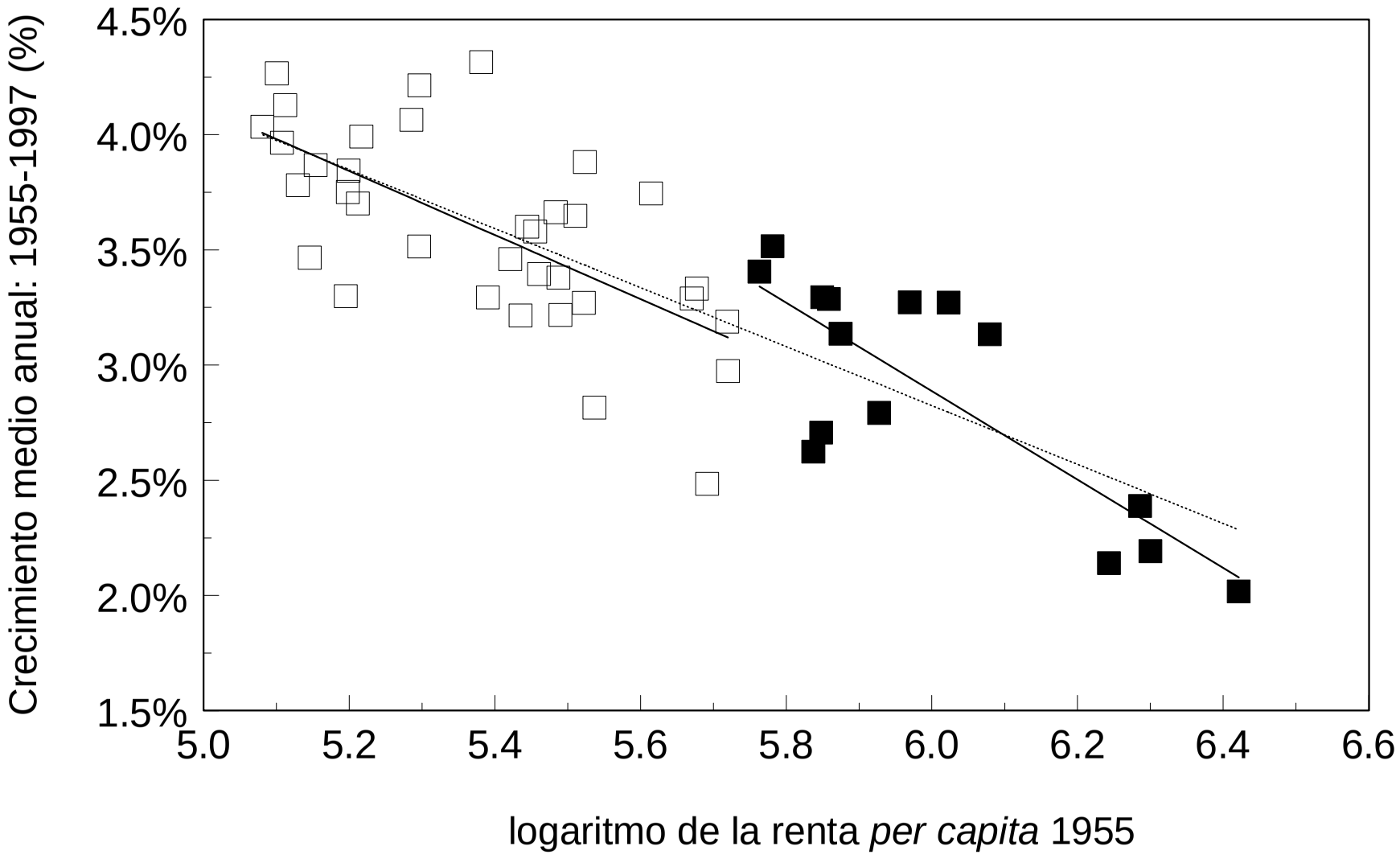
Gráfico 20

El examen de ambos grupos separadamente sugiere los siguientes **comentarios**. La velocidad de convergencia es ahora sustancialmente mayor dentro del grupo de provincias ricas que cuando son consideradas todas en conjunto, alcanzando ahora un valor del 3.90% para el periodo 1955-1997, de forma que dentro de este grupo las provincias con menor renta *per capita* en 1955 tendieron a crecer más que aquellas con un nivel de renta *per capita* mayor. Para las provincias pobres, las velocidad de convergencia es similar a cuando consideramos a todas las provincias conjuntamente, dicho valor es estimado ahora en un 2.09% para el conjunto de provincias pobres. Esta situación se ilustra claramente en el gráfico 20. Así pues el fenómeno descrito como β-convergencia es un fenómeno generalizado para ambos grupos de provincias, pero obsérvese que ahora se mide dicha convergencia respecto a un estado estacionario diferente, es decir al estimar la ecuación (52) para cada grupo suponemos implícitamente que el estado estacionario al que convergen los diferentes grupos de provincias es distinto. Observamos igualmente que la velocidad de convergencia no parece estable en el tiempo, siendo mayor en el periodo

[40] Para el periodo 1955-1997 los resultados son prácticamente idénticos a los del periodo 1955-1995. Las diferencias más relevantes se comentan en el texto.

Gráfico 20. *Beta*-convergencia: Ricos *versus* Pobres

Periodo: 1955-1997



1965-1985 que en las dos décadas inicial y final. Además la bondan del ajuste, que es similar o ligeramente mejor en el caso de las provincias ricas, empeora de forma sensible para el caso de las provincias pobres.

Las consideraciones anteriores hacen aconsejable contrastar la **hipótesis de estabilidad estructural** entre los coeficientes de la ecuación de convergencia para ambos grupos de provincias, **dicho contraste fue incapaz de rechazar la hipótesis nula de estabilidad**, al 5% de nivel de significación, tanto en la ecuación lineal (52) como en la no-lineal (46), e independientemente de utilizar inferencia robusta frente a heterocedasticidad o inferencia estándar. El mismo resultado se obtuvo al contrastar la igualdad de coeficientes por separado, así como un contraste de igualdad en los estados estacionarios de ambos grupos de provincias, el intercepto *a* en (49) o (52), condicionado en la igualdad en la velocidad de convergencia. En consecuencia la ecuación (53) puede considerarse una buena aproximación al conjunto de provincias y el resto de ejercicios de esta sección serán realizados sin distinguir entre ambos grupos.

Comentario 5: b-*convergencia* versus s-*convergencia.*

Ya hemos indicado como el concepto de **b-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. De hecho esto es lo que sucede en nuestra muestra, acabamos de ver como entre 1955 y 1997 obtenemos una velocidad de β-convergencia del 1.83% anual y al mismo tiempo el coeficiente de variación[41] pasó de 0.39 en 1955 a 0.21 en 1997, idéndicos resultados se obtienen utilizando la desviación típica del logaritmo que pasa de 0.35 a 0.21 entre ambos años. Sin embargo es bien conocido que esta relación no tiene

[41] En términos de estadísticos simples, pero el mismo fenómeno es observado en términos de estadísticos ponderados.

porque cumplirse y la existencia de **b-convergencia es compatible con diferentes situaciones en términos de s-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*[42] y tomando varianzas a ambos lados de (43) obtenemos

$$\sigma_t^2 = \rho^2 \sigma_{t-1}^2 + \sigma_u^2 \tag{54}$$

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 (54) obtenemos

$$\sigma_t^2 = \rho^{2j} \sigma_{t-j}^2 + \sigma_u^2 \Sigma_{k=0}^{j-1} \rho^{2k} \tag{55}$$

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

---

[42] Encontramos ahora otra razón para la popularidad de $Var(\log x_{it})$ como medida de dispersión *cross-section*, ciertamente en el contexto de (43) esta parece ser una medida natural si bien ya hemos observado en la sección anterior 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.

$$\sigma^2 = \lim_{j\to\infty}\sigma_t^2 = \frac{\sigma_u^2}{1-\rho^2} \tag{56}$$

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.

Por tanto 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 **s-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 **s-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**[43]. 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**[44].

El conjunto de gráficos 20 (a)-(c) permite ilustrar las diversas situaciones una vez eliminada la tendencia creciente en el estado estacionario debido a la existencia de progreso técnico (comentario 4), todas ellas partiendo de la misma situación de β-convergencia, $0<\beta<1 \Rightarrow 0<\rho<1$. En el gráfico 20(a) las economías parten de una

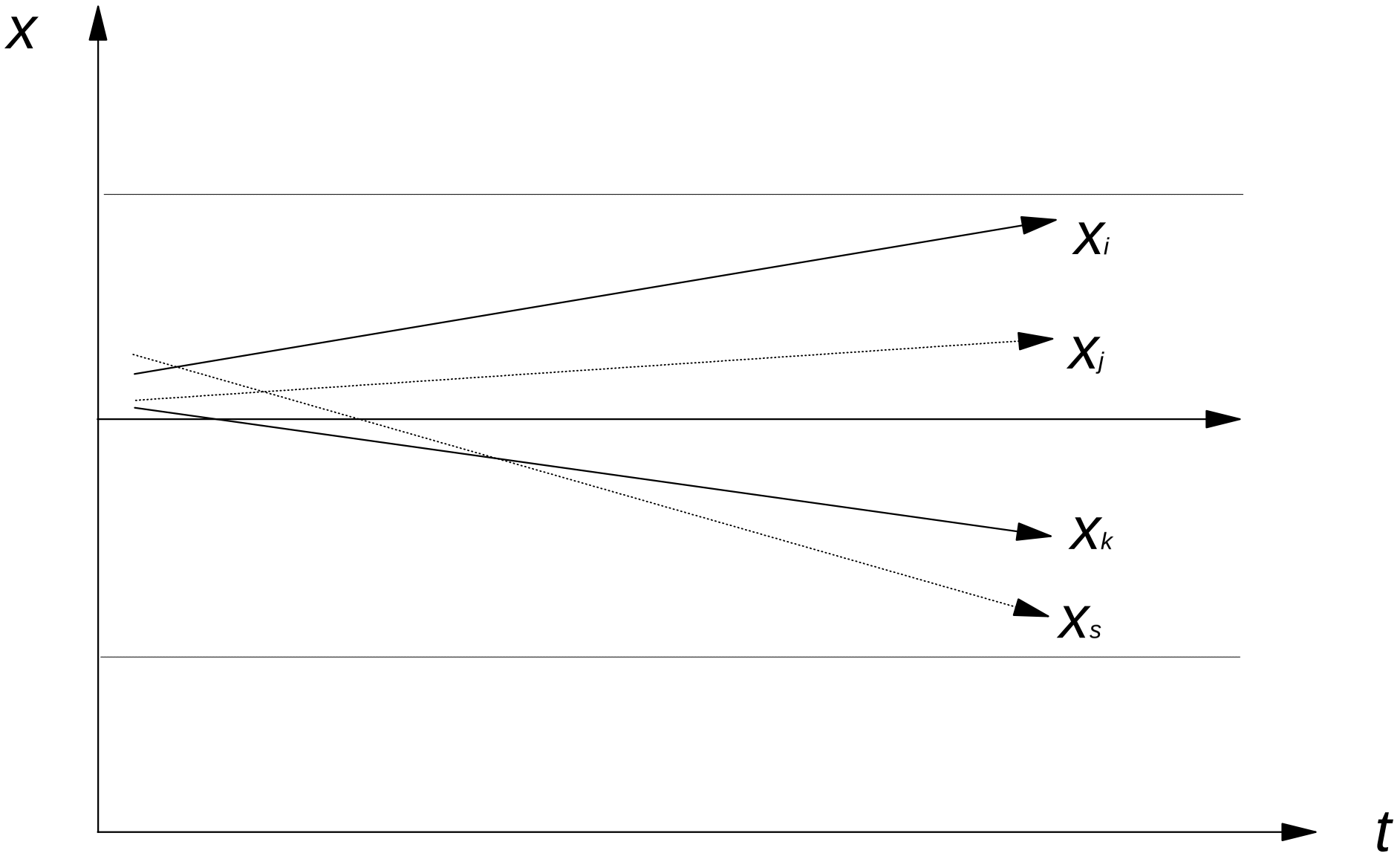
[43] Es necesario recordar que estamos razonando en términos poblacionales, en términos muestrales debemos observar aproximádamente estas características de la población.
[44] Obsérvese que β-divergencia implica necesariamente un incremento de la dispersión, lo que justifica la necesidad de β-convergencia para obtener σ-convergencia.

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 20(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. Esto parece ser, sin embargo, lo ocurrido con la renta *per capita* de las provincias españolas en el periodo 1955-1997.

Gráficos 21a,b,c,d

Finalmente el gráfico 21 (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 $1<\beta<2 \Rightarrow -1<\rho<0$. Por último el gráfico 21 (d) ilustra una situación similar a la del gráfico 20(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 21 (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 21 (c) y (d) es idéntica la situación es en realidad muy diferente. Son estos razonamientos los que permiten argumentar a Quah (1993a, 1993b, 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 las dinámica de las distribuciones en el tiempo. Las regresiones *cross-section* representan el comportamiento medio de un conjunto

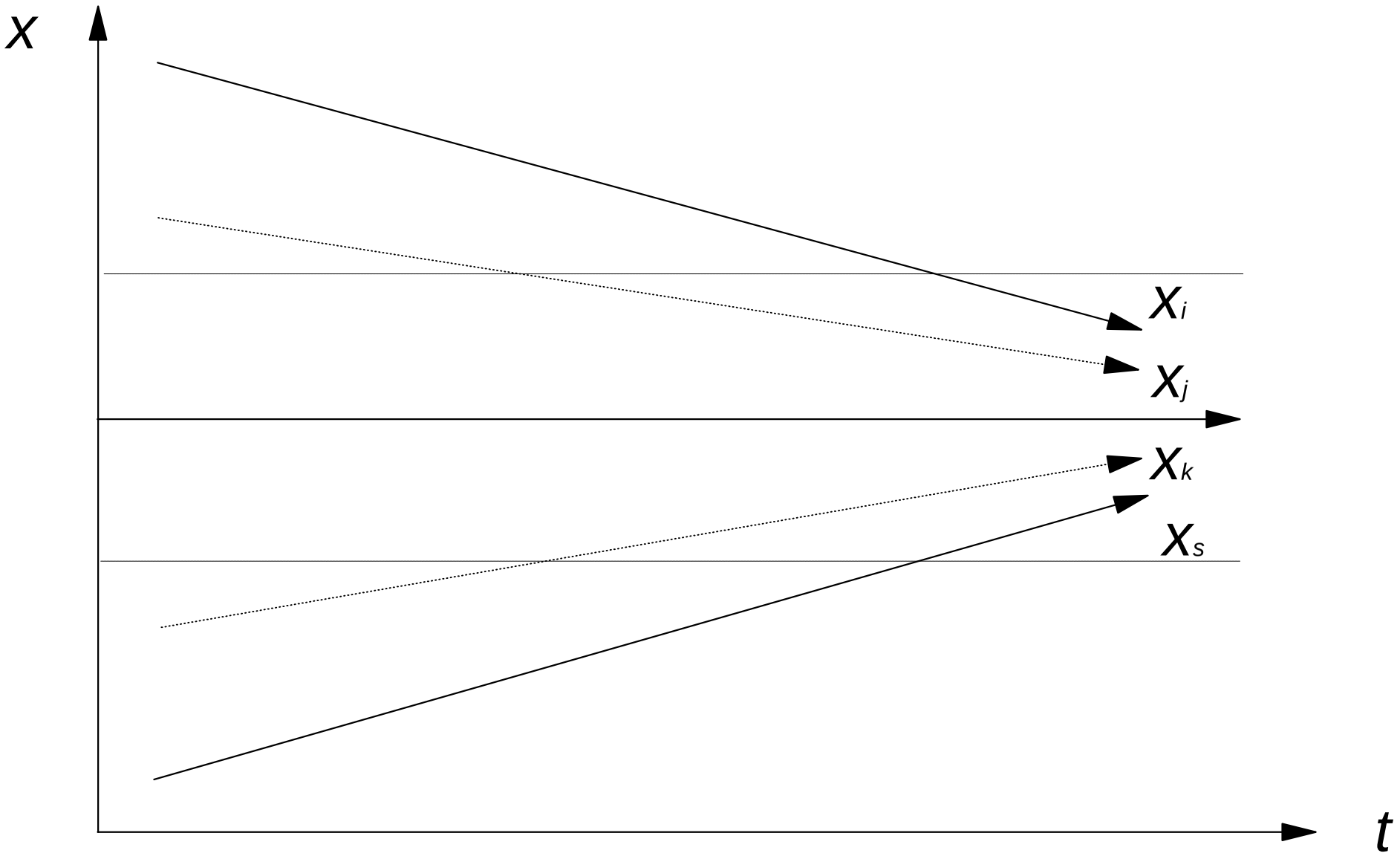
Gráfico 21 (a). *Sigma*-divergencia



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.

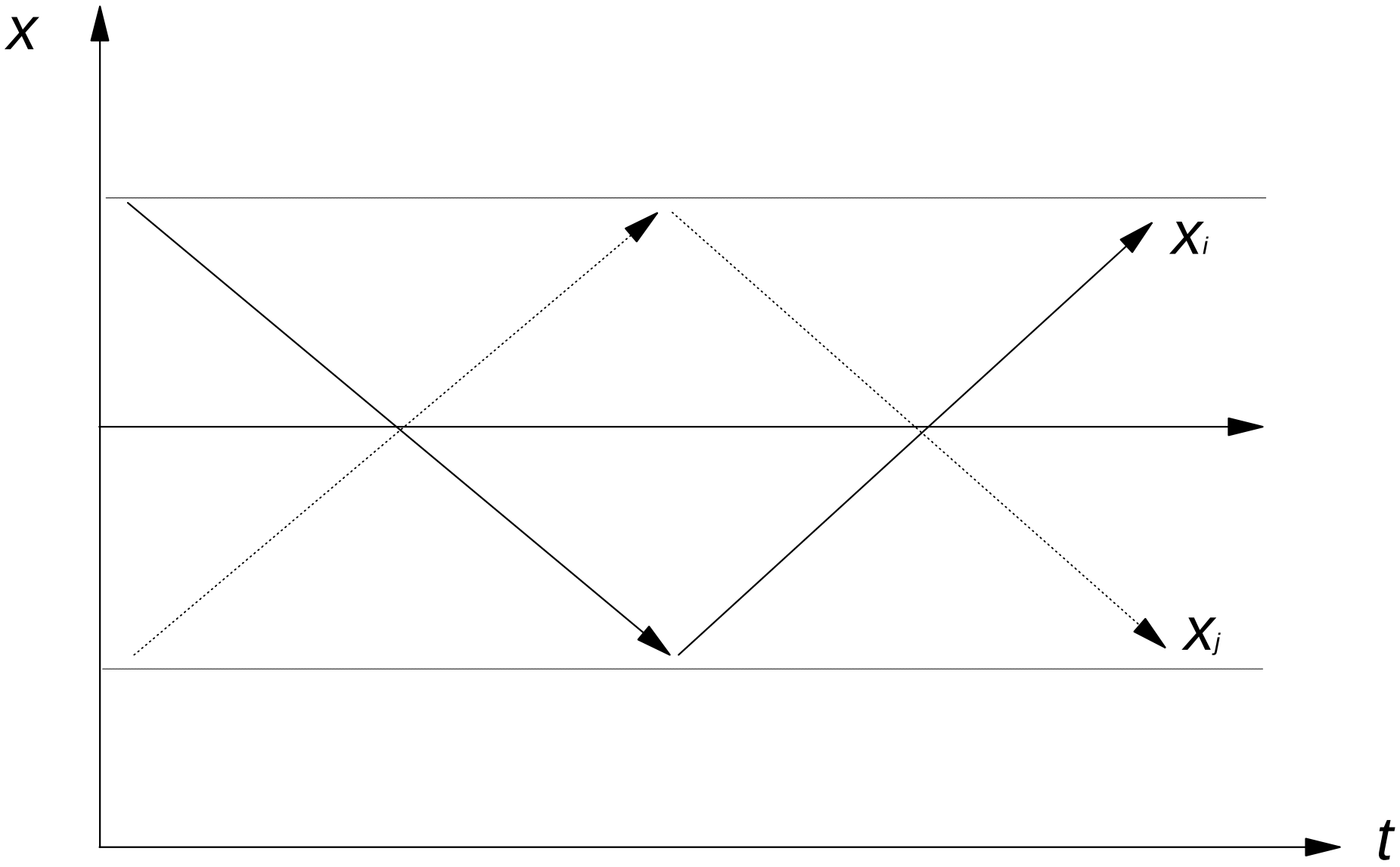
Gráfico 21 (b). *Sigma*-convergencia



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.

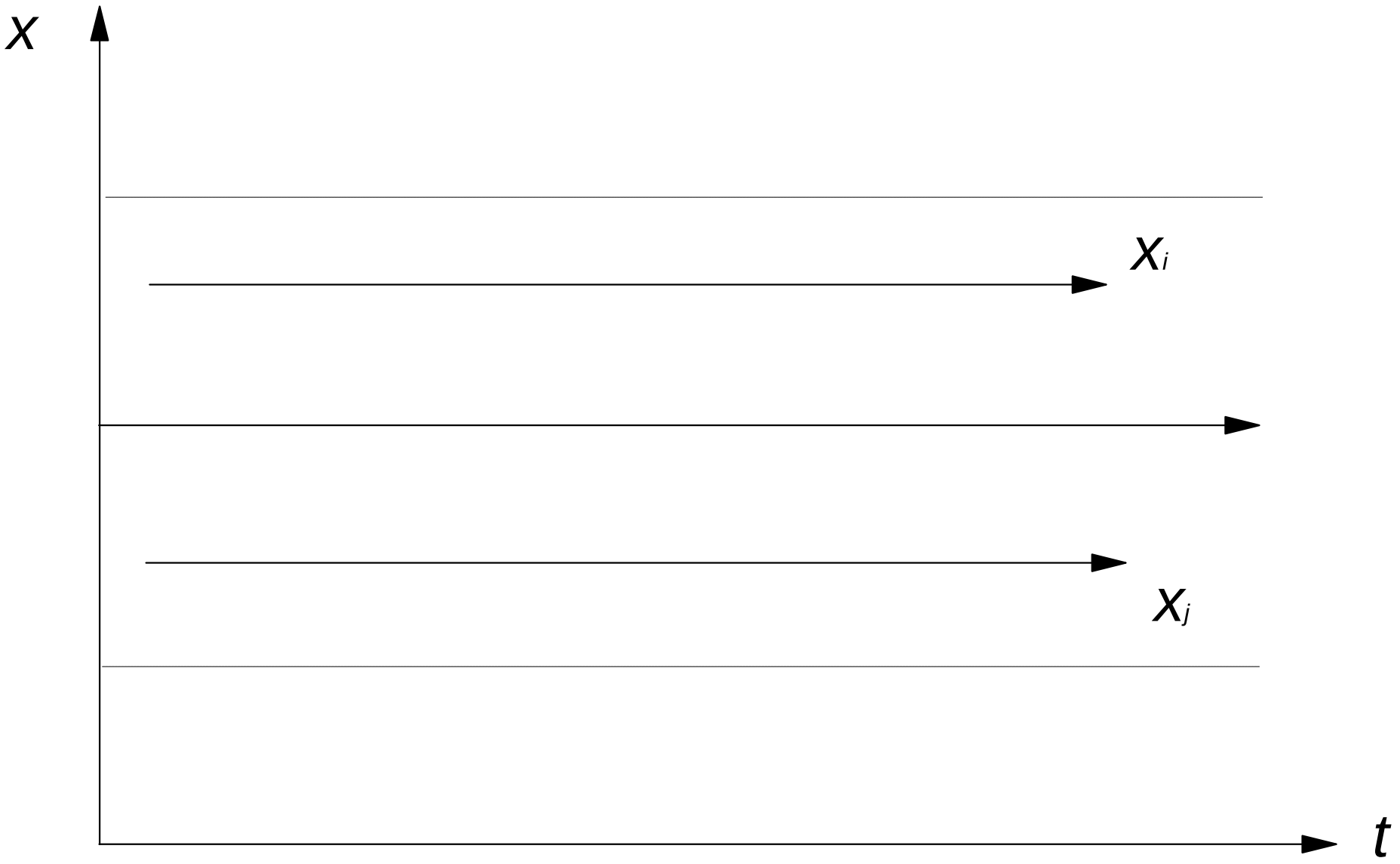
Gráfico 21 (c). *Sigma*-constante y gran movilidad



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 21 (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.

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 (56) es importante en el razonamiento anterior, desde el punto de vista de la teoría del crecimiento u_{it} se interpreta como perturbaciones transitorias, si estas no existieran 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* en 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 20.

La **relación entre** los modernos conceptos de **b-convergencia y s-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[45] 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 (43) puede ser escrita como

[45] 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.

$$E(y_{it}|y_{i,t-1}) = \mu + \rho(y_{i,t-1} - \mu) \tag{57}$$

siendo $E(y_{it}) = E(y_{i,t-1}) = \mu \;\; \forall t$ por estacionariedad[46] 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

$$\vartheta^2 = Cor^2(y_{it}, y_{i,t-1}) = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var(y_{i,t-1})} \leq 1 \tag{58}$$

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

$$\vartheta^2 = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var^2(y_{it})} = \rho^2 \leq 1 \quad \Rightarrow \quad |\rho| < 1 \;\text{ si }\; |\vartheta| < 1 \tag{59}$$

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 pero con tal que la correlación entre la situación inicial y final no sea idénticamente igual a 1 (gráfico 21 (d)) obtendremos β-convergencia[47].

Este es el argumento que permite a Quah (1993a) con razón 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

[46] α en (43) es por tanto $\alpha = \mu(1-\rho)$ en (57).
[47] Tampoco obtendríamos β-convergencia si la correlación entre la situación inicial y final fuera –1.

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** b**-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 (2000a)), 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 20(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[48].

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 **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, $E(y_{it}) = \mu_i \quad \forall t$, de forma que (57) se convierte en

$$E(y_{it} | y_{i,t-1}) = \mu_i + \rho(y_{i,t-1} - \mu_i) \tag{60}$$

[48] 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.

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

Podemos igualmente permitir un mayor grado de heterogeneidad en el corte transversal[49], 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**; este 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, $\vartheta^2 = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var(y_{i,t-1})}$, y del coeficiente de β-convergencia en (43), $\rho = \frac{Cov(y_{it}, y_{i,t-1})}{Var(y_{i,t-1})}$, es fácil obtener la siguiente relación

$$\vartheta^2 = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var(y_{i,t-1})} = \frac{Var(y_{i,t-1}).Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var^2(y_{i,t-1})} = \frac{Var(y_{i,t-1})}{Var(y_{it})}.\rho^2 \qquad (61)$$

por tanto,

$$\frac{Var(y_{it})}{Var(y_{i,t-1})} = \frac{\rho^2}{\vartheta^2} \qquad (62)$$

[49] 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.

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 de 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**. Así por ejemplo si para las provincias españolas tomamos como condición inicial 1955 y como condición final 1997 la correlación entre ambos periodos para el logaritmo de la renta *per capita* es de $\hat{\vartheta} = 0.78$, bastante elevada, lo que indica que la movilidad entre provincias en términos de renta *per*

capita es bastante reducida e intuitivamente debemos esperar que β-convergencia y σ-convergencia coincidan. Esto es en realidad lo que sucede en la práctica, ya que la estimación de mínimos cuadrados de ρ en (43) tomando 1955 como *t*–1 y 1997 como periodo *t* es notablemente inferior a dicho coeficiente de correlación siendo tan solo de 0.46[50], por lo que en nuestro ejemplo nos situamos en el caso (3).

Esto explica **porque β-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 σ-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 podíamos 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).

[50] Esta estimación se obtiene a partir de (53) como 1 – 42x0.012791.

Comentario 6: 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 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)[51]. 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[52] y dicha noción fue inicialmente denominada **reversión**. Sin embargo el trabajo que estableció la regresión

---

[51] 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 o acción de retroceder o volver hacia atrás (Anscombe (1967)).

[52] 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[53], 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 tambien 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)

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

[53] 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.

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 (57) y (58) 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[54].

Comentario 7: ¿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-convergencia** es un concepto dinámico que **relaciona la situación final con el crecimiento previo de una variable**.

[54] 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.

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

$$g_{x_i} = -\alpha' + \gamma x_{it} - u'_{i,t-1} \tag{63}$$

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 en el periodo t-1 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 (63)**, 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 (63), 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

$$\log x_{it} - \log x_{i,t-1} = -\alpha' + \gamma \log x_{it} - u'_{i,t-1} \tag{64}$$

que puede ser convenientemente escrita como

$$\begin{aligned} \log x_{i,t-1} &= \alpha' + (1-\gamma)\log x_{it} + u'_{i,t-1} \\ &= \alpha' + \rho' \log x_{it} + u'_{i,t-1} \qquad\qquad \rho' = 1-\gamma \end{aligned} \tag{65}$$

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 (65), 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 (65) contiene la misma información que (43), o alternativamente (63) la misma que (29). 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, aunque proporcionan informaciones diferentes**.

En el estudio sobre las tasas de crecimiento entre economías **¿debemos condicionar en el pasado o en el futuro?**[55], es decir ¿la ecuación de convergencia debe realizarse sobre la condición inicial, (29), tal y cómo normalmente se hace, o sobre la condición final, (63)?. 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 (comentario 5), resulta indiferente estimar la ecuación (60) o alternativamente

$$E(y_{i,t-1}|y_{it}) = \mu_i + \rho'(y_{it} - \mu_i) \qquad (66)$$

[55] La misma pregunta puede realizarse respecto a cualquier punto intermedio entre *t*–1 y *t* (Quah (1993a)).

donde $\rho' = \dfrac{Cov(y_{it}, y_{i,t-1})}{Var(y_{it})}$. Es fácil observar que $\rho\rho' = \dfrac{Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var(y_{i,t-1})} = \vartheta^2 \leq 1$, lo que clarifica la **relación entre los conceptos de** b**-convergencia y** g**-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 (29), β-convergencia, o (63), 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 (43) 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 (65) 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 a 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

$$\vartheta^2 = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var^2(y_{it})} = \rho'^2 \leq 1 \quad \Rightarrow \quad |\rho'| < 1 \text{ si } |\vartheta| < 1 \tag{67}$$

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[56].

[56] 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, $\vartheta^2 = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var(y_{i,t-1})}$, y del coeficiente de γ-convergencia en (65), $\rho' = \frac{Cov(y_{it}, y_{i,t-1})}{Var(y_{it})}$, es fácil obtener la siguiente relación

$$\vartheta^2 = \frac{Cov^2(y_{it}, y_{i,t-1})}{Var(y_{it}).Var(y_{i,t-1})} = \frac{Var(y_{it}).Cov^2(y_{it}, y_{i,t-1})}{Var^2(y_{it}).Var(y_{i,t-1})} = \frac{Var(y_{it})}{Var(y_{i,t-1})}.\rho'^2 \qquad (68)$$

por tanto,

$$\frac{Var(y_{i,t-1})}{Var(y_{it})} = \frac{\rho'^2}{\vartheta^2} \qquad (69)$$

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

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

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

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

Para el periodo 1955-1997 y el conjunto de provincias españolas la ecuación lineal de convergencia que condiciona en el peridodo final,

$$\frac{1}{j}\left(\log x_{it} - \log x_{i,t-j}\right) = -a' + c' \log x_{it} - \varepsilon'_{i,t-1} \tag{70}$$

produjo los siguientes resultados

$$\frac{1}{42}\left(\log x_{i,1997} - \log x_{i,1955}\right) = \underset{(.023879)}{.083334} - \underset{(.003441)}{.007095} \log x_{i,1997} - \hat{\varepsilon}'_{i,1997} \qquad R^2 = 0.0735 \tag{71}$$

donde entre paréntesis figuran los errores estándar consistentes frente a heterocedasticidad de forma desconocida (White (1980)) y el R^2 muestra un grado de ajuste bajo, 7.35%.

Sin embargo lo más importante es que esta ecuación muestra γ-divergencia al ser $\hat{c}' < 0$. En efecto, la estimación de mínimos cuadrados ordinarios de ρ' en (65) tomando 1955 como *t*–1 y 1997 como periodo *t* es 1.30[57], superior a la unidad, lo que produce una

[57] Esta estimación se obtiene a partir de (71) como 1 + 42x0.007095.

estimación de γ negativa, y en consecuencia tiende a mostrar que, por **términos medio**, **las provincias finalmente ricas no han crecido más que las provincias finalmente pobres**, sino todo lo contrario. En estas condiciones sabemos que la dispersión habrá disminuido, que es lo que efectivamente ha pasado, por lo tanto en nuestro ejemplo nos situamos en el caso (1), γ-divergencia y en consecuencia σ-convergencia. Obsérvese que $1.30 > 0.46$, que es la condición necesaria y suficiente para obtener σ-convergencia. El gráfico 22 ilustra de forma visual la ecuación estimada (71).

Gráfico 22

b-*convergencia: Datos de panel.*

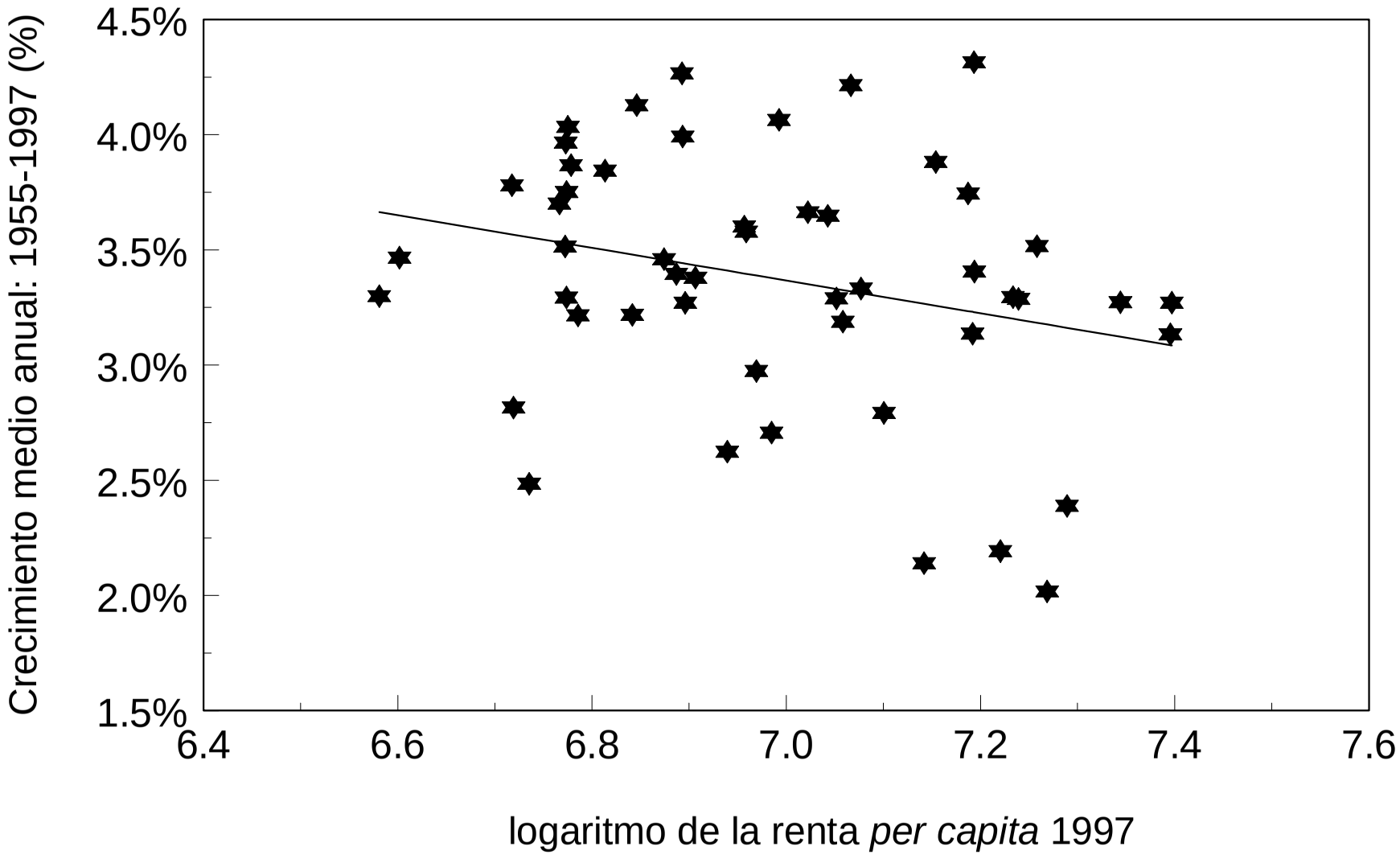
Excepto por los comentarios realizados al principio de esta sección acerca del concepto de β-convergencia en un contexto de series temporales y el contraste de raíces unidad con datos de panel toda nuestra exposición y aplicaciones prácticas se han centrado en la comparación de dos puntos del tiempo,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

Cuando la muestra se extiende más allá de dos peridos temporales, como es nuestro caso, 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 dimensiones, 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

Gráfico 22. *Gamma*-convergencia

Periodo: 1955-1997



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 se consideradas (Quah y Sargent (1993), Quah (1994a), Pesaran y Smith (1995)). En el contexto de modelos de regresión y dada la estructura de nuestro problema, que podemos considerar representado por la ecuación (42) 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 (1992)). 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), 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 (42), 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 (42) 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 3. Por las mismas razones expuestas en dicha sección y 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 (42) 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 (42) **efectos fijos individuales** (provinciales) 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. 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

[58] La posibilidad de recoger heterogeneidad observable por medio de otras variables explicativas será analizada en Goerlich (2000b).

[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).

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* como (53) 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** b**-convergencia en presencia de efectos fijos individuales significativos en realidad estamos argumentando a favor de la existencia de** b**-convergencia de cada economía a un estado estacionario diferente, lo que ciertamente desvirtúa el concepto de** b**-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** b**-convergencia presentan un *trade-off*** que cualquier analista deberá ponderar adecuadamente en presencia de una muestra concreta.

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

$$\log x_{it} - \log x_{i,t-1} = \alpha + \lambda_i + \eta_t - \beta \log x_{i,t-1} + u_{it} \qquad (72)$$

donde λ_i y η_t representan los efectos fijos individuales y temporales, tienen la misma interpretación que en (1) 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 presentar las estimaciones para nuestros datos conviene realizar **tres observaciones** de interés.

- **Primero**, desde el punto de vista de la **dinámica** ya hemos observado como la ecuación (72) 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 el análisis 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)).

  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)). Adicionalmente ya observamos al realizar el contraste de raíces unidad que la dinámica de nuestros datos parecía ser más general de lo que permite un solo desfase, lo que puede ser relevante en nuestro caso.

- **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 (72) se convierte en

$$\log z_{it} - \log z_{i,t-1} = \alpha + \lambda_i - \beta \log z_{i,t-1} + u_{it} \qquad (73)$$

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 (72) y (73) no son numéricamente equivalentes, salvo en el caso improbable en que

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

Que ecuación, (72) ó (73), 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 (72), 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.

  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 (2000b). Mayor heterogeneidad todavía es posible, de hecho algunos autores han argumentado que la heterogeneidad en las muestras utilizadas habitualmente en el

[60] Puesto que la media de agregado en cada año, μ_t, es una media ponderada y la regresión (73) 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).

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 (52) para datos de panel, es decir

$$\frac{1}{j}\left(\log x_{it} - \log x_{i,t-j}\right) = a + \lambda_i + \eta_t - b' \log x_{i,t-j} + \varepsilon_{it} \qquad (74)$$

donde los términos λ_i y η_t son idénticos a los de (72).

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

[61] El procedimiento es similar al propuesto para el contrastde de raíces unidad en el contexto de paneles heterogéneos (Im, Pesaran y Shin (1995), Maddala y Wu (1999)).

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 (49) 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 (1999), 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)).

En nuestro caso y como comparación de resultados estimaremos la ecuación (74) para el periodo 1955-1997 con periodicidad bianual, la máxima posible, y al mismo tiempo para el periodo 1955-1995 con periodicidad decenal, al igual que hicimos con el análisis de varianza de la sección 3.

Con periodicidad bianual, 1955-1997, la estimación de (74) **por mínimos cuadrados ordinarios** generó los siguientes resultados,

$$\frac{1}{2}\left(\log x_{it} - \log x_{i,t-2}\right) = \hat{\alpha} + \hat{\lambda}_i + \hat{\eta}_t - .041091 \log x_{i,t-2} + \hat{\varepsilon}_{it} \qquad R^2 = 0.7411 \tag{75}$$

(.005061)

Observaciones: 1050

			Nivel de significación
Efectos fijos provinciales:	F(49,979) =	2.49	0.0000
	CHI(49) =	159.93	0.0000
Efectos fijos temporales:	F(20,979) =	102.51	0.0000
	CHI(20) =	2691.83	0.0000

donde entre paréntesis figura el error estándar consistente frente a heterocedasticidad de forma desconocida (White (1980)), muy similar al error estándar habitual cuando se estima por mínimos cuadrados ordinarios que resultó ser de 0.005338, lo que es indicativo de que la heterocedasticidad no constituye un problema importante. En ambos casos la condición inicial resulta ser **altamente significativa**. El R^2 muestra un grado de ajuste sensiblemente superior al de la ecuación de convergencia *cross-section* (53) aunque sólo ligeramente más elevado que el del análisis de varianza del cuadro 2 que excluye $\log x_{i,t-2}$ como variable explicativa adicional[62]. Puede observarse además como ambos efectos fijos son altamente significativos.

La **velocidad de convergencia** anual implícita en la estimación de $\hat{b}' = 0.0411$ en (75) es del $\hat{b} = 4.29\%$, más del doble que la obtenida a partir de la ecuación *cross-section* (53) lo que está en linea con los trabajos de datos de panel que 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 ambos trabajos permitiendo un mayor grado de heterogeneidad, y De la Fuente (1998b) alrededor del 8% anual para una muestra y un periodo temporal similar al nuestro. Nuestra estimación se sitúa, por tanto, en el límite inferior de la amplia gama de estimaciones disponibles en la

[62] Existen pequeñas diferencias entre el análisis de varianza de la sección 3, donde las tasas de crecimiento se calculan como tasas medias anuales acumulativas, y la ecuación (74), donde las tasas de crecimiento se calculan como aproximaciones logarítmicas. Este pequeño cambio no altera los resultados del análisis de varianza de la sección 3 de forma que estos son comparables con los de la estimación en (75).

literatura aplicada con datos de panel[63]. En el contexto del modelo neoclásico con una velocidad de convergencia anual del 4% una economía tardaría algo más de 17 años en cerrar la mitad del *gap* existente entre el logaritmo de la renta *per capita* actual y el correspondiente al estado estacionario.

El **valor de** ρ **implícito** en esta estimación es de $\hat{\rho} = 0.96$. Dicho valor se reduce a 0.92 si tenemos en cuenta el caracter bianual de los datos y por tanto es muy similar al obtenido en el contraste de raíz unidad (48), lo que no es de extrañar ya que esta ecuación y (75) son enteramente equivalentes excepto por la mayor dinámica incluida en (48). **Aumentar la frecuencia de los datos, aumenta** pues **la velocidad de convergencia** y en consecuencia **disminuye la persistencia** respecto a la estimación *cross-section*.

De esta forma **la estimación con datos de panel refuerza la** β**-convergencia entre las provincias españolas** ya detectada anteriormente, pero **ahora a estados estacionarios diferentes** que si resultan ser ampliamente significativos. En consecuencia **no podemos concluir que las economías más pobres sean las más elejadas del estado estacionario y alternativamente las economías más ricas sean las más cercanas al estado estacionario** y **por tanto el análisis no arroja ninguna luz sobre la cuestión de si las provincias pobres han tendido a dar alcance (*cathing-up*) a las economías ricas en términos de renta *per capita***.

La ecuación estimada si nos permite, por contra, 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[64]. A partir de la versión determinista de (74) obtenemos que, en el estado estacionario, la renta *per capita* de la provincia *i*, x_i^*, puede obtenerse como

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

[63] Obsérvese además que esta velocidad es similar a la obtenida para las provincias ricas y la totalidad del periodo (cuadro 16).

[64] La estimación de (73) nos generaría la renta *per capita* del estado estacionario en términos relativos respecto a la media del agregado (De la Fuente (1998b)), que normalmente es creciente en el tiempo.

Estos valores y los efectos fijos estimados correspondientes a (75) se ofrecen en el cuadro 17 ordenados de mayor a menor. A grandes rasgos el *ranking* derivado de esta estimación se corresponde con la ordenación en términos de renta *per capita* observada (cuadro A2), de forma que aunque la distribución de la renta *per capita* provincial no haya alcanzado su distribución estacionaria la predicción de nuestro ejercicio es que las posiciones relativas de las distintas provincias no sufrirán grandes cambios en el futuro. Conviene recordar no obstante que estas estimaciones del estado estacionario no parece que sean muy fiables en términos de las rentas de equilibrio de largo plazo (De la Fuente (1998b)).

Cuadro 17

Con periodicidad decenal, 1955-1995, la estimación de (74) por **mínimos cuadrados ordinarios** generó unos resultados muy similares,

$$\frac{1}{10}\left(\log x_{it} - \log x_{i,t-10}\right) = \hat{\alpha} + \hat{\lambda}_i + \hat{\eta}_t - \underset{(.005180)}{.036658} \log x_{i,t-10} + \hat{\varepsilon}_{it} \qquad R^2 = 0.8987 \tag{76}$$

Observaciones: 200

			Nivel de significación
Efectos fijos provinciales:	F(49,146) =	2.29	0.0001
	CHI(49) =	307.43	0.0000
Efectos fijos temporales:	F(3,146) =	148.76	0.0000
	CHI(3) =	608.73	0.0000

donde entre paréntesis figura el error estándar consistente frente a heterocedasticidad de forma desconocida (White (1980)), de nuevo muy similar al error estándar habitual cuando se estima por mínimos cuadrados ordinarios que resultó ser de 0.005546, lo que vuelve a mostrar que la heterocedasticidad no constituye un problema importante. En ambos casos la condición inicial sigue resultando **altamente significativa**. El R^2 muestra un grado de ajuste sensiblemente superior al de la ecuación de convergencia *cross-section* del cuadro 16 aunque, al igual que en caso de los datos bianuales, sólo ligeramente más elevado que el del análisis de varianza del cuadro 2 que excluye $\log x_{i,t\text{-}10}$ como variable explicativa

Cuadro 17: Efectos fijos y renta per capita de estado estacionario

	Periodo Bienal	1955-1997 Efecto fijo	Renta *per capita*		Periodo Decenal	1955-1995 Efecto fijo	Renta *per capita*
13	Baleares	0.018032	2,031.223	13	Baleares	0.016546	1,904.882
47	Alava	0.017041	1,982.791	47	Alava	0.016404	1,897.504
32	Gerona	0.016268	1,945.855	32	Gerona	0.014942	1,823.341
34	Tarragona	0.013009	1,797.474	34	Tarragona	0.012394	1,700.915
29	Guadalajara	0.011595	1,736.703	33	Lérida	0.010308	1,606.803
33	Lérida	0.011577	1,735.930	29	Guadalajara	0.010150	1,599.919
50	La Rioja	0.010168	1,677.409	50	La Rioja	0.009287	1,562.694
46	Navarra	0.009566	1,653.016	46	Navarra	0.009169	1,557.671
31	Barcelona	0.008470	1,609.516	31	Barcelona	0.008227	1,518.151
36	Castellón	0.007488	1,571.505	44	Madrid	0.007872	1,503.499
18	Burgos	0.007458	1,570.357	36	Castellón	0.006927	1,465.247
44	Madrid	0.007430	1,569.267	18	Burgos	0.006786	1,459.629
11	Zaragoza	0.006924	1,550.084	11	Zaragoza	0.006271	1,439.265
9	Huesca	0.006429	1,531.522	9	Huesca	0.006159	1,434.856
10	Teruel	0.005481	1,496.578	48	Guipúzcoa	0.005328	1,402.687
15	SC de Tenerife	0.004515	1,461.806	49	Vizcaya	0.004866	1,385.135
49	Vizcaya	0.004104	1,447.259	15	SC de Tenerife	0.003678	1,340.980
48	Guipúzcoa	0.003844	1,438.145	10	Teruel	0.003654	1,340.102
24	Valladolid	0.003457	1,424.671	24	Valladolid	0.003468	1,333.321
14	Las Palmas	0.003057	1,410.846	37	Valencia	0.002750	1,307.436
37	Valencia	0.002866	1,404.298	14	Las Palmas	0.002721	1,306.423
20	Palencia	0.001946	1,373.229	20	Palencia	0.001605	1,267.233
35	Alicante	0.000612	1,329.358	35	Alicante	0.001321	1,257.477
23	Soria	0.000580	1,328.318	16	Cantabria	0.000599	1,232.941
16	Cantabria	-0.000152	1,304.867	23	Soria	0.000471	1,228.628
22	Segovia	-0.001275	1,269.704	12	Asturias	-0.000167	1,207.454
30	Toledo	-0.001478	1,263.445	22	Segovia	-0.001121	1,176.414
12	Asturias	-0.001558	1,260.987	45	Murcia	-0.001624	1,160.393
21	Salamanca	-0.001700	1,256.627	30	Toledo	-0.001767	1,155.885
45	Murcia	-0.002338	1,237.270	21	Salamanca	-0.002097	1,145.527
40	La Coruña	-0.003023	1,216.806	40	La Coruña	-0.002472	1,133.847
43	Pontevedra	-0.003390	1,205.979	43	Pontevedra	-0.002679	1,127.486
39	Cáceres	-0.004647	1,169.669	5	Huelva	-0.004202	1,081.594
5	Huelva	-0.005204	1,153.917	27	Ciudad Real	-0.004624	1,069.200
27	Ciudad Real	-0.005235	1,153.040	19	León	-0.004938	1,060.099
19	León	-0.005518	1,145.125	7	Málaga	-0.005232	1,051.634
17	Avila	-0.006011	1,131.472	17	Avila	-0.005755	1,036.726
7	Málaga	-0.006079	1,129.598	39	Cáceres	-0.006002	1,029.768
28	Cuenca	-0.007669	1,086.723	1	Almería	-0.007427	990.502
1	Almería	-0.007847	1,082.019	28	Cuenca	-0.008117	972.038
26	Albacete	-0.008918	1,054.187	41	Lugo	-0.008226	969.154
41	Lugo	-0.009104	1,049.428	26	Albacete	-0.008461	962.958
25	Zamora	-0.009704	1,034.205	42	Orense	-0.008878	952.065
42	Orense	-0.009852	1,030.492	25	Zamora	-0.008917	951.043
6	Jaén	-0.010362	1,017.798	2	Cádiz	-0.010719	905.433
2	Cádiz	-0.011933	979.613	8	Sevilla	-0.011082	896.519
3	Córdoba	-0.011972	978.693	6	Jaén	-0.012766	856.252
8	Sevilla	-0.012753	960.247	3	Córdoba	-0.012881	853.567
4	Granada	-0.016908	867.899	4	Granada	-0.015163	802.067
38	Badajoz	-0.017287	859.929	38	Badajoz	-0.016591	771.411
	Media	0.000000	1339.538		Media	0.000000	1243.915
	Desviación típica	0.008702	287.238		Desviación típica	0.008229	281.176

adicional. Puede observarse además como ambos efectos fijos siguen siendo altamente significativos.

La **velocidad de convergencia** anual implícita en la estimación de $\hat{b}' = 0.0367$ en (76) es del $\hat{b} = 4.57\%$, muy similar a la obtenida con datos bianuales, aunque ligeramente más elevada en esta última estimación, y mucho mayor que la obtenida a partir de la ecuación *cross-section* del cuadro 16. El **valor de** ρ **implícito** en (76) es de $\hat{\rho} = 0.96$.

Por tanto a pesar del diferente lapso temporal entre las ecuaciones (75) y (76) los resultados son ampliamente similares y la consideración de un desfase temporal u otro no altera las conclusiones esenciales, incluidas las referentes a la estimación de las rentas *per capita* de estado estacionario correspondientes a (76) y que tambien se ofrecen en el cuadro 17. En consecuencia podemos afirmar de nuevo que **la estimación con datos de panel refuerza la** β**-convergencia entre las provincias españolas**, aunque **ahora a estados estacionarios diferentes**.

El principal **problema** con la estimación de la **ecuación (74)** 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 (72) en forma autoregresiva

$$\log x_{it} = \alpha + \lambda_i + \eta_t + \rho \log x_{i,t-1} + u_{it} \qquad \rho = 1 - \beta \tag{77}$$

cuya única diferencia con (43) 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**[65] **individuales** en **modelos dinámicos** de **datos de panel** provoca que el **estimador** de **mínimos cuadrados ordinarios** en (77) sea **inconsistente** cuando $n \to \infty$ y $\boldsymbol{T}$ **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 (77) sea independiente e idénticamente distribuido, tanto

[65] La constante, α, es suficiente para generar el resultado que mencionaremos a continuación.

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 (77) si es consistente y asintóticamente equivalente al estimador de máxima verosimilitud bajo normalidad[66] (Amemiya (1967)).

Sabemos además que cuando T es pequeño y $\rho > 0$, como es nuestro caso, entonces el sesgo de mínimos cuadrados ordinarios es negativo (Hsiao (1986), Sec.-4.2). En consecuencia, puesto que $\beta = 1 - \rho$, es probable que las **estimaciones de la velocidad de convergencia derivadas de (75) y (76) estén sesgadas al alza**. Ciertamente para el caso de las observaciones bianuales $T = 22$, un valor no excesivamente pequeño, y además los resultados con un lapso temporal decenal fueron muy similares de forma que quizá en nuestro caso los sesgos no sean excesivamente grandes (De la Fuente (1998b), p.-10), pero conviene asegurarse de ello examinando la robustez de las estimaciones anteriores mediante la comparación de los resultados ya obtenidos con un estimador que sea consistente aún cuando T sea fijo. De hecho 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.

[66] 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 tambien inconsistente mientras que T sea fijo y esta inconsistencia sólo desaparece cuando $T \to \infty$.

La estimación consistente de (77), o alternativamente (72), 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 (77) 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

$$\log x_{it} - \log x_{i,t-1} = (\eta_t - \eta_{t-1}) + \rho(\log x_{i,t-1} - \log x_{i,t-2}) + (u_{it} - u_{i,t-1}) \tag{78}$$

o

$$\Delta \log x_{it} = \Delta\eta_t + \rho\Delta \log x_{i,t-1} + \Delta u_{it} \tag{79}$$

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

En términos de β (79) puede ser escrito como

$$\Delta \log x_{it} - \Delta \log x_{i,t-1} = \Delta\eta_t - \beta\Delta \log x_{i,t-1} + \Delta u_{it} \tag{80}$$

y dado que el lapso temporal entre observaciones no suele ser igual a un periodo la ecuación que se estima es análoga a (74)

$$\frac{1}{j}\left(\Delta \log x_{it} - \Delta \log x_{i,t-j}\right) = \Delta\eta_t - b'\Delta \log x_{i,t-j} + \Delta\varepsilon_{it} \tag{81}$$

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 (81) 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 (78) por mínimos cuadrados ordinarios no es apropiada**[67]. 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 (78) 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

$$\mathrm{E}\left[\left(\Delta \log x_{it} - \Delta\eta_t - \rho\Delta \log x_{i,t-1}\right)\log x_{i,t-s}\right] = 0 \qquad s = 2,\ldots,(t-1);\ t = 3,\ldots,T \qquad (82)$$

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

Considerando a (78) como un sistema hipotético de $T-2$ ecuaciones podemos construir a partir de (82) un **estimador generalizado de momentos** de ρ que sea óptimo dentro de su 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

[67] 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 (78) por mínimos cuadrados ordinarios sería adecuada (Arellano y Bond (1991), p.-282).

fijos temporales, obtenemos la siguiente **asignación dinámica de instrumentos** para cada ecuación

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

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 matríz de covarianzas del término de perturbación en (77) las que permiten la estimación de los parámetros de interés (Ahn y Schmidt (1990)). 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

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

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

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

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

en consecuencia la matríz de varianzas-covarianzas del término de perturbación en (78) es conocida hasta un factor de proporcionalidad.

De la misma forma que hicimos con el modelo (1) de la sección 3 podemos formular el modelo (77) en **notación de muestra completa** suponiendo una **organización de las observaciones por individuo**, al igual que en la sección 4. Teniendo en cuenta que el carácter dinámico de (77) nos hace perder una observación temporal por cada individuo podemos escribir

$$\log \mathbf{x} = \ell_{n(T-1)}\alpha + \mathbf{D}_n \mathrm{l} + \mathbf{D}_{T-1}\mathrm{h} + \log \mathbf{x}_{-1}\rho + \mathbf{u} \tag{83}$$

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 matríz $n(T-1)$xn que contiene el conjunto de las n variables ficticias individuales, l es un vector nx1 de efectos fijos individuales, $\mathbf{D}_{T-1} = \ell_n \otimes \mathbf{I}_{T-1}$ es una matríz $n(T-1)$x$(T-1)$ que contiene el conjunto de las $T-1$ variables ficticias temporales, 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 matríz de diferenciación temporal para cada individuo, i

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

y para todo los individuos como

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

podemos escribir el modelo (78) en notación de muestra completa premultiplicando (83) por $\mathbf{A}$.

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 (78) puede ser expresado en notación de muestra completa como

$$\mathbf{A}\log\mathbf{x} = \mathbf{AD}_{T-1}\mathrm{h} + \mathbf{A}\log\mathbf{x}_{-1}\rho + \mathbf{Au} \tag{84}$$

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 (84) 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 (77) (Caselli, Esquivel y Lefort (1996)), o simplemente centrar nuestro análisis empírico en $z_{it} = \frac{x_{it}}{\mu_t}$ 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 (77) de la que hemos partido y la ecuación transformada (78) 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 (83) para el individuo i, $\mathbf{u}_i = (u_{i2}, u_{i3}, \ldots, u_{iT})'$, se obtiene que $\mathrm{Var}(\mathbf{A}_\Delta \mathbf{u}_i) = \sigma_u^2 \mathbf{A}_\Delta \mathbf{A}_\Delta' = \sigma_u^2 \mathbf{H}$, 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 $\mathrm{Var}(\mathbf{Au}) = \sigma_u^2 (\mathbf{I}_n \otimes \mathbf{H})$.

Así pues dada la matríz 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 (82), 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 (82) utilizados en el procesos de estimación. Para el individuo *i* la matríz 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 (84) dicha matríz viene dada por

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

De esta forma dado el sistema (84), la matríz de instrumentos **Z** y $\mathrm{Var}(\mathbf{Au})=\sigma_u^2(\mathbf{I}_n\otimes\mathbf{H})$, conocida hasta el factor de proporcionalidad s_u^2, es directo obtener un **estimador generalizado de momentos óptimo** bajo nuestras hipótesis hacerca de u_{it} utilizando como **matríz de ponderaciones** (Hansen (1982), White (1982))

$$\mathbf{W}_n=\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i'\mathbf{H}\mathbf{Z}_i\right)^{-1}=\left(\frac{1}{n}\mathbf{Z}'(\mathbf{I}_n\otimes\mathbf{H})\mathbf{Z}\right)^{-1} \qquad (85)$$

Este **estimador** puede ser calculado en **una sola etapa** y es **óptimo** bajo el **supuesto** de **que 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 (84).

Si **mantenemos** el supuesto de **independencia**[68] pero **permitimos** la existencia de **heterocedasticidad de forma desconocida**, **tanto en el corte transversal como en la dimensión temporal**, entonces el estimador que utiliza como matríz 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 (82) y en ausencia de hipótesis adicionales. Este es necesariamente un **estimador en dos etapas**.

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

Dados los resíduos de (84) 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 matríz 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 **matríz de ponderaciones**

$$\tilde{\mathbf{W}}_n = \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{Z}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \mathbf{Z}_i \right)^{-1} \tag{86}$$

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 (84) 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** $\tilde{\mathbf{W}}_n$ 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 dimensiones**, **dentro de la clase de estimadores basados en las restricciones de momentos lineales (82) 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 (84) 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 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 (78) y sustituyendo dichas estimaciones en (77) es posible obtener una estimación de $\alpha + \lambda_i + u_{it}$ como

$$\overline{(\alpha + \lambda_i + u_{it})} = \log x_{it} - (\hat{\eta}_t + \hat{\rho} \log x_{i,t-1}) \tag{87}$$

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

$$\widehat{(\alpha + \lambda_i)} = \frac{1}{T-1} \sum_{t=1}^{T-1} \overline{(\alpha + \lambda_i + u_{it})} \tag{88}$$

Antes de examinar los resultados derivados de utilizar el estimador de momentos que hemos descrito cuando es aplicado a nuestro conjunto de datos 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} tambien 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**. Por esta razón estas restricciones no serán utilizadas en la estimación.

- 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)), pero no serán consideradas en nuestro ejemplo (De la Fuente (1998b)).

- Puesto que la solución apuntada consiste en eliminar los efectos individuales, λ_i en (77), 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 (77) 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 $E(\lambda_i) = 0 \ \forall i$ 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*) nuestra muestra es exahustiva, provincias, 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 (77) 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 tambien 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 utilizamos en nuestros cálculos es potencialmente mejorable en términos de eficiencia. De hecho existe cierta evidencia de Monte Carlo de que dicho estimador, que utiliza sólo los momentos (82), 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))[69], pero sólo recientemente se han introducido en el contexto de paneles dinámicos (Islam (1995), Caselli, Esquivel y Lefort (1996), De la Fuente (1998b), Forbes (1998)).

El estimador generalizado de momentos que hemos descrito fue aplicado a nuestros datos con una periodicidad decenal para el periodo 1955 - 1995, $T = 5$. El motivo de no utilizarlo con un lapso de tiempo bianual responde a las siguientes razones. Primero, los

[69] 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)).

resultados de la estimación de (74) a partir de mínimos cuadrados ordinarios y con los dos lapsos de tiempo diferentes, estimaciones (75) y (76), fueron muy similares, por lo que tampoco sería de esperar ahora grandes diferencias entre las estimaciones correspondientes a ambas periodicidades. Segundo, si consideramos un lapso de tiempo bianual entonces en número de observaciones en la dimensión temporal es moderadamente elevado, $T = 22$. En estas circunstancias las inconsistencias derivadas del estimador de mínimos cuadrados ordinarios no deben ser muy grandes y este puede ser preferible a otros estimadores consistentes sobre la base de la minimización del error cuadrático medio asintótico como criterio de optimalidad. Tercero, y fundamentalmente, el número de restricciones de identificación derivadas de (82) son $m = (T-2).(T-1)/2$, lo que muestra que el estimador mencionado está pensado para paneles con n relativamente grande y T relativamente pequeño, ya que cuando $T = 22$ las restricciones de identificación en nuestro modelo son $m = 20 \text{x} 21/2 = 210$. ¡Y todas ellas para estimar un solo parámetro!. En estas circunstancias el modelo está altamente sobreidentificado y el comportamiento en muestras finitas del estimador puede no mostrar las ganancias de eficiencia que la teoría asintótica nos indica. Por ello el estimador de momentos no fue aplicado a los datos con un lapso de tiempo bianual a pesar de que una solución posible consistiría simplemente en descartar algunos instrumentos, tomando solamente aquellos más altamente correlacionados con las variables instrumentadas. Así, por ejemplo, en nuestro contexto podríamos utilizar un solo instrumento por ecuación, $m = 20$, que de forma natural debería ser el nivel desfasado dos periodos, $\log x_{i,t-2}$.

La estimación de (81) mediante el método generalizado de momentos con periodicidad decenal, 1955 - 1995, y utilizando $\mathbf{W}_n$ como matríz de ponderaciones generó los siguientes resultados,

$$\frac{1}{10}\left(\Delta \log x_{it} - \Delta \log x_{i,t-10}\right) = \Delta\hat{\eta}_t - \underset{(0.012250)}{0.013095}\,\Delta \log x_{i,t-10} + \Delta\hat{\varepsilon}_{it} \qquad \mathrm{R}^2 = 0.8338 \tag{89}$$

Observaciones: 150

		Nivel de significación
Restricciones de sobreidentificación:	CHI(5) = 7.86	0.1636

puesto que este es el **estimador en una etapa** asume implícitamente perturbaciones independiente e idénticamente distribuídas, tanto en el corte transversal como en el ámbito temporal.

La estimación de b' en (89) es sensiblemente inferior a la que obtuvimos por mínimos cuadrados ordinarios, (76), lo que refleja el sesgo al alza de esta última estimación. De hecho la **velocidad de convergencia** anual implícita en $\hat{b}' = 0.0131$ es del $\hat{b} = 1.40\%$, notablemente inferior a la obtenida a partir de la ecuación *cross-section* (52) (cuadro 16) y sorprendentemente baja para las estimaciones habituales en la literatura. En el contexto del modelo neoclásico a esta velocidad una economía tardaría casi 50 años en cerrar la mitad del *gap* existente entre el logaritmo de la renta *per capita* actual y el correspondiente al estado estacionario. Sin embargo la ecuación (81) representa una relación entre la aceleración del crecimiento y la tasa de crecimiento inicial de la renta *per capita*, de forma que ahora la intuición no es tan directa a pesar de las "teóricas" mejoras econométricas en las estimaciones. El grado de bondad del ajuste del modelo es aceptable, si bien el R^2 no tiene la interpretación habitual al tratarse (89) de una ecuación estimada por variables instrumentales. ¡El **valor de** $\mathbf{r}$ **implícito** en esta estimación es de $\hat{\rho} = 0.99$!.

A pesar de todo el resultado más llamativo de (89) es que **la estimación de** b' **no es significativa**, con una *t-ratio* de 1.07, lo que pone en entredicho todos los resultados anteriores respecto a la β-convergencia de las provincias españolas y entra en contradicción con el contraste de raíz unidad (48). Podríamos pensar que ello es debido a la falta de validez de los instrumentos, ya que este contraste detectó en los datos una dinámica más rica que la impuesta por la ecuación de convergencia, pero el contraste de Sargan (1958)-Hansen (1982) de restricciones de sobreidentificación nos permite descartar esta interpretación.

Alternativamente ya hemos visto como el estimador en una etapa es consistente pero no eficiente, sin embargo relajar el supuesto de homocedasticidad y construir el estimador en dos etapas, que utiliza $\tilde{\mathbf{W}}_n$ como matríz de ponderaciones en la segunda etapa, no mejora las cosas, sino más bien al contrario. La estimación de b' de dos etapas disminuye a aproximadamente la mitad, $\hat{b}' = 0.0067$, y sigue sin ser significativa, con una

t-ratio de 1.01. Esta estimación representa una velocidad de convergencia ridículamente baja, $\hat{b} = 0.69\%$. Sin embargo merece la pena observar que el contraste de restricciones de sobreidentificación presenta ahora un nivel de significación de 0.0391, lo que si indica un potencial problema con los instrumentos.

Todo ello nos induce a pensar que el estimador de momentos generalizado utilizado no presenta buen comportamiento en muestras finitas (Alonso-Borrego y Arellano (1996), Blundell y Bond (1998)). La poca precisión del estimador, junto con el descenso en la estimación puntual de la velocidad de convergencia, hacen aparecer a esta como no significativa e indican la conveniencia de examinar otros momentos adicionales en la estimación. En concreto sería deseable examinar la posibilidad de incorporar diferencias desfasadas de $\log x_{it}$, $\Delta \log x_{i,t-s} \quad s \geq 1$, como instrumentos adicionales para ecuaciones en niveles (Blundell y Bond (1998)), aunque ello supusiera una reinterpretación de los efectos individuales.

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.

Comentario final

Un **comentario final** respecto a los conceptos de β-convergencia y σ-convergencia. Muchos de los argumentos relacionados con estos conceptos tienen como referencia una **distribución unimodal**, de forma que por ejemplo una reducción de la dispersión puede ser entendida de forma intuitiva como una reducción de las desigualdades[70]. Sin embargo, todas estas ideas colapsan en **distribuciones multimodales**, donde los conceptos de

[70] Sólo de forma intuitiva, ya que habría que analizar las propiedades concretas del índice de dispersión utilizado así como el hecho de si las curvas de Lorenz se cruzan o no (Goerlich (1998)).

polarización (Esteban y Ray (1993, 1994), Esteban (1996)) y **“clubs”** (Baumol (1986), DeLong (1988)) se vueleven relevantes. Es por ello que ir más allá de los estadísticos de dispersión y prestar atención a otros estadísticos que nos ayuden a caracterizar la forma de la distribución de nuestra variable de interés, $\phi(x)$, como los analizados en la sección anterior, o mirar directamente la forma de dicha distribución por medio de los métodos que examinaremos en la sección siguiente es importante.