Convergence in Panel Data: Evidence from the Skipping Estimation

Etsuro Shioji*
Department of Economics and Business
Universitat Pompeu Fabra†

August 20, 1997

This paper demonstrates that, unlike what the conventional wisdom says, measurement error biases in panel data estimation of convergence using OLS with fixed effects are huge, not trivial. It does so by way of the "skipping estimation": taking data from every m years of the sample (where m is an integer greater than or equal to 2), as opposed to every single year. It is shown that the estimated speed of convergence from the OLS with fixed effects is biased upwards by as much as 7 to 15%.

*This research was partly financed by the DGICYT grant PB94-1095. I would like to thank Fabio Canova and Xavier Sala-i-Martin for stimulating discussions.

†From October 1997, Department of Economics, Yokohama National University, 79-3 Tokiwadai, Hodogaya-ku, Yokohama 240, Japan. e-mail: shioji@econ.ynu.ac.jp.

1 Introduction

Estimating the speed of convergence has been one of the central issues of recent empirical literature of economic growth. At some point, it seemed like there was a general consensus among economists on this issue: economies do converge to their respective steady states, but at a very slow rate. However, many of the recent studies cast doubt on this view and conclude that the speed of convergence is actually very fast. It is important to resolve this controversy because different speeds of convergence have very different implications for an economy that deviates from its steady state. On one hand, cross section estimation of speed of convergence generally yield the estimated speed of about two percent (Barro and Sala-i-Martin (1992), for example). This speed means a half life of deviation from one's steady state of about 34.3 years. On the other hand, panel data estimation of the speed, the approach that has been employed by many recent authors, almost always produce much higher estimates, ranging from 4 to 20% For example, De la Fuente (1996) applies OLS with Fixed Effects to the Spanish regional data and comes up with the estimated speed of about 12%. Canova and Marcet (1995) apply their Bayesian estimation method to the regional data from Western Europe and countries in Europe, and come up with the estimated speed of about 20% for the former and 10% for the latter (though the values vary depending on the precise specification). Islam (1995) applies OLS with Fixed Effects as well as the minimum distance estimation method to the Summers Heston data set and comes up with estimates of between 4% and 9.3 % depending on the method and countries included in the regression. Also, Casselli, Esquivel and Lefort (1996), Evans (1995), and Knight, Laoyza and Villanueva (1993) apply different methods of panel data estimation to the Summers Heston data set and come up with estimated speed of convergence much higher than 2%. A speed of 10% (which seems to be more or less the mean of these estimates) would mean a half life of deviation from one's own steady state of about 6.6 years. Different estimates have very different implications for growth theories as well. The neoclassical growth model of Solow (1956) and Swan (1956), when one takes a narrow view on capital stock and identifies it solely as physical capital, implies a speed of 5-6 % (Barro and Sala-i-Martin (1995)). Hence, if the results from cross section regressions are to be trusted, the definition of capital stock in the model must be extended to include human capital or some other types of capital. On the other hand, if the mean of the panel data estimates of about 10% is to be trusted, there is no way the neoclassical growth model alone can explain such a high speed of convergence, as the speed of 10% implies a negative share of capital if the neoclassical growth model is taken literally (De la Fuente (1996))[1].

This paper criticizes the panel data estimation of convergence with non-instrumental variables approach for being seriously biased. I will argue that the true speed of convergence is likely to be much closer to the ones found in the cross section regressions. My investigation will concentrate on OLS with Fixed Effects of De la Fuente (1996), as analytical results are easy to derive in this case. I plan to write a companion paper that deals with the method of Canova and Marcet (1995), but there I will have to rely heavily on Monte Carlo experiments. That these estimates are biased itself is well known. There are two types of biases. First, these estimates are subject to small sample biases. Second, they are subject to measurement error biases. They both tend to bias the estimated speed of convergence upwards. The extent of the latter type of biases depends on the standard deviation of the measurement error relative to that of the true shocks. It has customarily been argued that this type of bias cannot be too serious, as errors in measurement of income or output in developed countries cannot be so large compared to true shocks to the economy (Canova and Marcet (1995)).

However, this argument is not sufficient to reject the importance of the measurement error biases when we are talking about estimating long run tendencies of economies using annual panel data. A typical exercise in this literature is to estimate the following model:

$$y_{it} = (1 - \beta) \cdot y_{it-1} + \alpha_i + u_{it} \tag{1}$$

where y_{it} is output per capita of region i in year t relative to, say, the mean of the whole country, β is a parameter that is called the **speed of convergence**, that is common across regions and periods ($|1-\beta| < 1$), α_i is the region specific fixed effect term, u_{it} is shock to the true output per capita with mean 0 and variance σ_u^2 which is assumed to be serially and spatially independent. The main point is that this is a growth model. In other words, this model is meant to capture long run movements in output per capita. When Solow wrote his paper on growth, for example, he did not mean his model to be a representation of year-to-year behavior in output. Hence, y_{it} should be interpreted as long run output per capita, net of short run movements such as business cycle elements and temporary shocks (e.g. a bad harvest, a temporary increase in oil prices, effects of having an expo in one year and thus getting millions of tourists and zillions of governmental subsidies, etc.). The error term u_{it} should also be interpreted as representing shocks to the

[1]On the other hand, if one takes the lower bound estimate of 4%, this value is consistent with the neoclassical growth model with the narrow definition of capital stock.

long run level of output. It is true that, by taking deviations from the country means, this approach reduces the business cycle elements in regional output per capita to some extent, but, as regional business cycles are not perfectly synchronized, and as some regions tend to have more share of sectors that are sensitive to business cycles than others, it would not completely take out business cycle elements. From this viewpoint of growth theory, all these short run components in output per capita should be considered as "measurement errors" in a broad sense of the word. Once we take this broad view of measurement errors, it is not clear anymore if these errors can be considered small compared to true shocks.

The question is, then, how to know the extent of biases due to measurement errors in the broad sense. This paper proposes a simple approach based on the "skipping estimation". There is nothing fancy about this. All it means is that, instead of using the whole yearly data in output, we use data from every m years, where m is equal to or greater than two. It will be shown that, if the measurement error biases are trivial, this skipping should push the estimated speed of convergence upwards compared to the non-skipping estimate. This is because skipping worsens the small sample bias. Only when the measurement error biases are serious, skipping pushes the estimated speed of convergence downwards. This is because skipping lessens the measurement error bias. Hence, by looking at effects of skipping on the estimated speed of convergence, we can infer the extent of biases due to measurement errors (in the broad sense).

I present evidence from the US states, the Japanese prefectures, and OECD countries. In all the cases, it is shown that skipping lowers the estimated speed of convergence or at least has no sizable effect, for m not too large. Hence, these results suggest that the measurement error biases are serious, not trivial. It is shown that, in some cases, the measurement errors may be biasing the estimates upwards by as much as 20%.

Before I proceed to the main text, let me clarify what this paper does NOT say. The purpose of this paper is not to promote panel data estimation of convergence WITH instrumental variables. It is true that the instrumental variable estimator of Anderson and Hsiao (1981), which was employed by Evans, yields consistent estimates even in the presence of measurement errors if one takes enough number of lags in picking up instruments. However, I found that this method is way too inefficient to draw any meaningful conclusion: when I applied this method to the US states, the point estimate for the speed of convergence varied between -4.6% and 8.4%, its standard errors being between 13% and 19%. In other words, in none of the cases could I reject the null hypothesis of the true speed being zero. On the other hand, the GMM approach of Arellano and Bond (1991), the technique used by Caselli, Esquivel and Lefort (1996), yields a much more efficient estimate. The problem with this method is that the number of instruments needed increases very quickly as the sample size increases in the time dimension. This makes use of yearly data impossible in cases where data spans for more than 15 years or so, due to constraints on computer ability. One has to take data from every few years in order to avoid crushing a computer. In other words, the very nature of the method forces one to do the skipping.

The rest of the paper is organized as follows. Section 2 explains how the extent of the biases depends on various parameters. Section 3 explains how skipping affects these biases. Section 4 presents empirical results and argues that OLS with Fixed Effects produces serious biases. Section 5 concludes.

2 Biases in the OLS with Fixed Effects

It has been established that, even in the absence of measurement errors, that is, even if the model in equation (1) is a correct one, the OLS with fixed effect estimator for β, which will be denoted by $\hat{\beta}$ here, is subject to small sample bias. Let's say there are N regions ($i = 1, 2, ..., N$) and the sample last for T periods ($t = 1, 2, ..., T$). Then, as N goes to infinity with T fixed, $\hat{\beta}$ does not converge to the true β. Nerlove (1971) shows this using Monte Carlo experiments. Nickell (1981) derives the precise analytical expression for the bias:

$$\hat{\beta} - \beta = B_1/B_2$$

where

$$B_1 = \frac{1}{T^2} \cdot \frac{(T-1) - T \cdot (1-\beta) + (1-\beta)^T}{\beta^2}$$

$$B_2 = \frac{1}{1-(1-\beta)^2} \cdot \left\{ 1 - \frac{1}{T} - \frac{2 \cdot (1-\beta)}{\beta^2} \cdot \frac{(T-1) - T \cdot (1-\beta) + (1-\beta)^T}{T^2} \right\}.$$

These expressions are shown in Hsiao (1986). The bias is positive unless T is too small. Hence, the estimated speed is biased upwards. For a reasonably large value of T, the above bias is approximately

$$\hat{\beta} - \beta \approx (2-\beta)/(T-1)$$

(Nickell (1981)). Note that, even if we are fortunate enough to have T as large as 100 (which is almost unheard of in the literature of convergence regressions), for a reasonable value of β (say between 0.001 and 0.1), the upward bias is about two percent[2]. If what we are interested in were the value of the AR(1) coefficient, $1-\beta$, this bias would not look that big. But here what we are truly interested in is the value of β and thus the bias of two percent could have a significant consequence. For example, if the true speed of convergence is two percent as many economists who work on cross section regressions claim, then, the bias of two percent would mean that the estimated speed of convergence would on average be double the true speed. The two percent convergence implies a half life of deviation from steady state of 34.3 years. The four percent convergence implies a half life of 17.0 years. Thus the economic interpretation of the result would be substantially different. As T goes to infinity, this bias disappears.

When the measurement error is present, the model has to be rewritten in the following way. Let y_{it} be the observed output per capita of region i in period t, relative to the country mean in that period. Assume that y_{it} can be described by the following model:

$$y^*_{it} = (1-\beta)\cdot y^*_{it-1} + \alpha_i + u_{it}$$

and

$$y_{it} = y^*_{it} + v_{it}$$

where y^*_{it} is the "true" long run output per capita and v_{it} is the measurement error with mean 0 and variance σ_v^2 which is also assumed to be serially and spatially independent. Distributions of both u_{it} and v_{it} are assumed to be space- and time-invariant. The model in equation (1) is equivalent to the above model with $v_{it} = 0$ for all i and t. If the measurement errors are present, it can be shown that the above expression for the asymptotic bias for $N = +\infty$ is modified to:

$$\hat{\beta} - \beta = (B_1 + B_3)/(B_2 + B_4)$$

where

$$B_3 = \left((1-\beta) - (1-\beta)\cdot\frac{1}{T} + \frac{T-1}{T^2}\right)\cdot\left(\frac{\sigma_v}{\sigma_u}\right)^2$$

[2]A beauty of the small sample bias is that it is not very sensitive to the true value of β. Therefore, as long as one can assume that the true β falls in some "reasonable" range, a likely range of values for this bias can be narrowed down fairly well. This, unfortunately, is not the case for the measurement error bias.

and

$$B_4 = \left(1 + \frac{1}{T}\right) \cdot \left(\frac{\sigma_v}{\sigma_u}\right)^2.$$

Note that B_3 is positive for any $T > 1$, and B_4 is always positive. Hence, the whole bias is still positive. It now depends on three things: the true β, T, and the ratio between the standard deviation of measurement errors and that of the true shock, σ_v/σ_u. To measure the contribution of measurement errors to the whole bias, note that, as stated before,

$$B_1/B_2 \approx (2 - \beta)/(T - 1)$$

while

$$B_3/B_4 \approx 1 - \beta$$

hence, for a reasonably large T, $B_1/B_2 < B_3/B_4$. It follows that the presence of measurement errors tends to worsen the upward bias in the OLS with fixed effects estimator for the speed of convergence.

The question is how important the worsening of the bias is. Table 1 shows the asymptotic bias for different values of T and σ_v/σ_u. Note that these are biases for the speed of convergence, β, rather than the AR(1) coefficient $1-\beta$. Hence, a positive value means that the estimated speed is biased upwards, and therefore that $1 - \beta$ is biased downwards. Moreover, the numbers are in percentages. Table 1A corresponds to the case where the true speed β is two percent and in Table 1B the true speed β is set to 10%. First, look at the first column in each of the tables. In this column, the standard deviation of measurement errors is set to zero. Hence, in this case, only the small sample bias exists. It can be seen that the bias is severe even for high values of T such as 60 or 120. Next, look at the last row in each table. Here, T is set to $+\infty$. Hence, in this case, there is no small sample bias and only the measurement error bias exists. Note that the measurement error bias really kicks in only when the ratio σ_v/σ_u is fairly large. For $\beta = 2\%$, the bias becomes large only for the ratio greater than 0.5. For $\beta = 10\%$, the bias becomes large for the ratio much greater than 0.2. As pure measurement errors are expected to be much smaller than the true shock term, if we take the narrow view of measurement errors, the bias cannot be that important. This is the reason why the measurement error bias has not been considered as a big problem in the literature. But if we take the broad view of measurement errors, values such as $\sigma_v/\sigma_u = 0.5$ or even 1 are quite possible. And in such a case, bias due to these errors could be serious. Moving to the other rows of the table, it can be seen that the measurement error bias is even more serious when the sample size is finite. For example, compare the last row of Table 1A ($T = +\infty$) with the fourth row ($T = 60$). In the last row, in the absence of the small sample bias, increasing the ratio σ_v/σ_u from 0 to 0.5 increases the upward bias in the speed of convergence by less than 1 percent. However, in the fourth row, even if T is as large as 60, increasing the ratio σ_v/σ_u from 0 to 0.5 increases the upward bias in the speed of convergence by almost 3 percent. Hence, there is no simple relationship between the small sample bias and the measurement error bias such as (total bias) = (small sample bias) + (measurement error bias). On the contrary, **when the sample size is finite, the two types of bias reinforce each other** to often create a huge bias. For example, for the case where the true speed is 2%, $T = 60$ and $\sigma_v/\sigma_u = 0.50$, the total bias is as large as 7.40%, so the estimated speed of convergence will be on average 9.40%. Hence it is important to know the extent of measurement errors, to pin down the likely amount of bias.

Table 1: Biases in Estimated Speed of Convergence: $\hat{\beta} - \beta$ in percentage

A: When the True Speed β is 2 %

T\\$\frac{\sigma_v}{\sigma_u}$	0	0.1	0.2	0.5	1	2
10	26.65	27.07	28.30	35.69	51.77	71.89
20	13.74	14.01	14.81	19.99	33.96	59.60
40	6.83	6.99	7.47	10.70	20.52	44.27
60	4.47	4.59	4.95	7.40	15.16	36.16
120	2.11	2.19	2.42	4.03	9.34	25.66
1000	0.21	0.25	0.38	1.26	4.30	14.76
$+\infty$	0.00	0.04	0.15	0.96	3.73	13.40

B: When the True Speed β is 10 %

T\\$\frac{\sigma_v}{\sigma_u}$	0	0.1	0.2	0.5	1	2
10	24.32	24.76	26.03	33.58	49.29	67.69
20	11.96	12.29	13.27	19.42	34.86	59.45
40	5.63	5.88	6.63	11.49	24.94	50.98
60	3.60	3.82	4.49	8.86	21.34	47.24
120	1.70	1.90	2.48	6.35	17.75	43.14
1000	0.19	0.37	0.88	4.34	14.76	39.38
$+\infty$	0.00	0.17	0.68	4.08	14.37	38.86

3 the Skipping Estimation

This paper tries to investigate the extent of the measurement error bias using the skipping estimation. The idea is to use the data from every m years, instead of every single years. Hence, the estimated equation is

$$y_{it} = (1-\beta)^m \cdot y_{it-m} + \left[1 + (1-\beta) + \cdots + (1-\beta)^{m-1}\right] \cdot \alpha_i + e_{mit}$$

where

$$e_{mit} = u_{it} + (1-\beta) \cdot u_{it-1} + \cdots + (1-\beta)^{m-1} \cdot u_{it-m+1}.$$

This model can be estimated with standard OLS with fixed effects:

$$y_{it} = \gamma_m \cdot y_{it-m} + \alpha_{im} + e_{mit} \tag{2}$$

where γ_m is an empirical counterpart of $(1-\beta)^m$. Let its estimated value be $\hat{\gamma}_m$. Then I recover estimated β by the formula

$$\hat{\beta}_m = 1 - \hat{\gamma}_m^{1/m}.$$

The bias of this estimator can be derived analytically. Appendix A presents this derivation. The bias is computed numerically under various different assumptions in Table 2 and Table 3. Table 2 deals with the case where $T = +\infty$, or the case when there is no small sample bias. Table 2A shows the case in which the true speed of convergence is 2%, and Table 2B shows the case in which the speed is 10%. Different rows correspond to different values of m. The values of m is chosen for the sake of comparison with the later results. It is clear that, in this case, the bias decreases with m. In other words, **skipping reduces the measurement error bias**.

Table 2: Biases in the Skipping Estimation
When T is infinite(in percentage)
A: When the True Speed β is 2 %

m$\backslash\frac{\sigma_v}{\sigma_u}$	0	0.1	0.2	0.5	1	2
1	0.00	0.04	0.15	0.96	3.73	13.40
2	0.00	0.04	0.15	0.94	3.66	13.13
3	0.00	0.04	0.15	0.92	3.59	12.87
4	0.00	0.04	0.15	0.90	3.51	12.61
5	0.00	0.04	0.14	0.89	3.44	12.36
6	0.00	0.04	0.14	0.87	3.37	12.11
10	0.00	0.03	0.13	0.80	3.11	11.17
12	0.00	0.03	0.12	0.77	2.99	10.73
15	0.00	0.03	0.12	0.72	2.81	10.10
20	0.00	0.03	0.11	0.65	2.54	9.13

B: When the True Speed β is 10 %

m$\backslash\frac{\sigma_v}{\sigma_u}$	0	0.1	0.2	0.5	1	2
1	0.00	0.17	0.68	4.08	14.37	38.86
2	0.00	0.15	0.61	3.67	12.93	34.98
3	0.00	0.14	0.55	3.31	11.64	31.48
4	0.00	0.12	0.49	2.98	10.48	28.33
5	0.00	0.11	0.45	2.68	9.43	25.50
6	0.00	0.10	0.40	2.41	8.49	22.95
10	0.00	0.07	0.26	1.58	5.57	15.06
12	0.00	0.05	0.21	1.28	4.51	12.20
15	0.00	0.04	0.16	0.93	3.29	8.89
20	0.00	0.02	0.09	0.55	1.94	5.25

Table 3 deals with the case where T is finite. Table 3A and 3B correspond to the case when T is 60, and Table 3C and 3D correspond to the case when T is 40. In Table 3A and 3C, the true speed of convergence is set to 2%, and in Table 3B and 3D it is 10%. Note, first, that, when there is no measurement error (the first column in each panel), the bias worsens with m. That is, **skipping worsens the small sample bias**. When both the small sample bias and the measurement error bias are present, the result becomes a mixture of the two extreme cases. When measurement errors are unimportant, say when σ_v/σ_u is 0.1 (the second column in each panel), skipping always worsens the bias. The bias-correcting effect of skipping starts to appear as σ_v/σ_u increases. With $\sigma_v/\sigma_u = 0.2$ (the third column), there is a minor reduction in the bias only when m is increased from 1 to 2. As m is increased further, the bias worsens. Note that, when the true β is high (namely when it is 10%) the bias worsens drastically for a large m. With $\sigma_v/\sigma_u = 0.5$ (the fourth column), there is a sizable improvement in the bias till m reaches 4 or 5 and then the bias starts to worsen. Again , the worsening is drastic when β is large. With $\sigma_v/\sigma_u = 1.0$ (the fifth column), the bias reaches its trough when m is much larger (10 in Table 3A, 6 in Table 3B, 8 in Table 3C and 5 in Table 3D). The conclusion is that, **if measurement errors are really unimportant, skipping should worsen the bias, not improve it.** Hence, by studying if skipping increases the estimated β (a symptom of worsening bias) or decreases it (a symptom of improving bias), one could know the extent of the measurement error bias.

Table 3: Biases in the Skipping Estimation

(in percentage)

A: $T = 60$, $\beta = 2\%$

m$\backslash\frac{\sigma_v}{\sigma_u}$	0	0.1	0.2	0.5	1	2
1	4.47	4.59	4.95	7.40	15.16	36.16
2	4.56	4.62	4.80	6.03	10.04	21.88
3	4.66	4.70	4.82	5.64	8.33	16.46
4	4.77	4.80	4.89	5.50	7.51	13.64
5	4.89	4.91	4.98	5.47	7.06	11.94
6	5.01	5.03	5.09	5.49	6.81	10.81
10	5.67	5.68	5.71	5.93	6.64	8.72

B: $T = 60$, $\beta = 10\%$

m$\backslash\frac{\sigma_v}{\sigma_u}$	0	0.1	0.2	0.5	1	2
1	3.60	3.82	4.49	8.86	21.34	47.24
2	3.86	3.97	4.30	6.53	13.21	29.26
3	4.16	4.24	4.46	5.96	10.53	21.97
4	4.52	4.58	4.75	5.88	9.37	18.25
5	4.96	5.00	5.14	6.06	8.89	16.19
6	5.49	5.52	5.64	6.42	8.84	15.09
10	9.83	9.85	9.94	10.55	12.45	17.97

C: $T = 40$, $\beta = 2\%$

m$\backslash\frac{\sigma_v}{\sigma_u}$	0	0.1	0.2	0.5	1	2
1	6.83	6.99	7.47	10.70	20.52	44.27
2	7.01	7.09	7.33	8.95	14.05	27.77
4	7.43	7.47	7.58	8.37	10.88	17.80
5	7.68	7.71	7.80	8.41	10.36	15.68
8	8.68	8.69	8.74	9.05	10.03	12.52

D: $T = 40$, $\beta = 10\%$

m$\backslash\frac{\sigma_v}{\sigma_u}$	0	0.1	0.2	0.5	1	2
1	5.63	5.88	6.63	11.49	24.94	50.98
2	6.08	6.20	6.58	9.05	16.28	32.58
4	7.30	7.36	7.55	8.80	12.52	21.35
5	8.15	8.20	8.35	9.36	12.36	19.47
8	13.30	13.33	13.43	14.09	16.02	20.61

4 Evidence from the Skipping Estimation

Evidence from the US states

Skipping requires a reasonably long series. Perhaps the longest existing series that has been used in the literature of convergence is the US state per capita personal income. BEA has collected data every year from 1929. The latest year for which the data was available at the time of this research was 1996. This means that there is data for 68 years, for the 48 contiguous state. Using the values for 1929 as the initial values (y_{i0}), we are still left with $T = 67$ years of data. Thus the US states data seems to be the most ideal place to start with. Table 4 shows the estimated speed of convergence β from OLS with Fixed Effects estimation for different values of m. Unfortunately, 67 is not as nice a number as 60: it cannot be divided by too many integers. Hence, the sample size is different for different values of m. Define T_m as the largest integer that is equal to or smaller than 67 and that can be divided by m. Then T_m is the sample size for that particular value of m. The second column shows this sample size, and the third column shows the implied last period of the sample. For example, for $m = 2$, the largest integer that is smaller than 67 that can be divided by m is 66, so the sample period starts from 1929 (for $t = 0$) and ends in 1995 ($t = 66$). The fourth column is the estimated speed of convergence β in percentage, and the last column is the standard error of the estimate, also in percentage (see Appendix B for the formula). All the data sources used in this paper are summarized in Appendix C.

The table shows that, when $m = 1$, that is, when there is no skipping, the estimated speed of convergence is 8.6%. This result of fast convergence is comparable to the findings of De la Fuente (1996) and Canova and Marcet (1995). The question is if this is because β is truly large, or it simply reflects a bias that is inflated by the presence of measurement errors. Going down the rows, the table shows the general tendency for the estimated β to decrease with m for smaller values of m and then to rise with m for larger m's. That is, the bias improves as skipping becomes larger, at least for smaller m's. This could not happen if measurement errors were truly negligible. In fact, the estimate reaches its bottom at $m = 11$, which suggests that measurement errors are actually quite large. In Table 3 in the previous section, Panel A (in which T is set to 60) shows that, if the true speed is 2%, the bias reaches its minimum at $m = 10$ only when the ratio σ_v/σ_u is as large as 1.0. In such a case, Table 3A indicates that the non-skipping estimate of the speed (the case where $m = 1$) is biased by as much as 15%. Even if we take a conservative stance and assume that σ_v/σ_u is 0.5, the non-skipping bias is still over 7%. Hence, the estimated speed of convergence is subject to a serious upward bias.

Table 4: Evidence from the US States

the estimated speed is in percentage

m	T_m	end year	$\hat{\beta}_m$	std.
1	67	96	8.64	0.67
2	66	95	7.48	0.63
3	66	95	6.67	0.59
4	64	93	7.33	0.64
5	65	94	6.84	0.61
6	66	95	6.20	0.57
7	63	92	6.02	0.58
8	64	93	5.73	0.53
9	63	92	6.37	0.61
10	60	89	5.21	0.55
11	66	95	3.90	0.44
12	60	89	4.72	0.52
13	65	94	5.07	0.52
14	56	85	5.28	0.58
15	60	89	7.06	0.68
16	64	93	6.36	0.58

Evidence from the Japanese prefectures

Data for the Japanese per capita Prefectural Income is available for 1950-1990. As I was particularly concerned with the quality of the 1950 data (see Appendix C), I use the 1951 values for initial values ($t = 0$). There are currently 47 prefectures in Japan but data on Okinawa is not available for earlier years as it was under US occupation till 1972. Hence the sample size is $T = 39$ and $N = 46$. The result is summarized in Table 5 in the same way as in Table 4. First, note from the first row that the non-skipping estimate suggests a very high speed of convergence of 16.9% per year. However, going down the rows, there is a broad tendency for the estimate to decrease with m (with a notable exception of the case $m = 2$) for smaller m and then to increase with m. The bottom is reached at $m = 6$. Table 3C and 3D (in which T is set to 40) suggest that this can happen only if the ratio σ_v/σ_u is greater than 0.5. Moreover, the reduction in the estimated speed due to skipping is very large: between $m = 1$ and $m = 6$, the estimated speed goes down by more than 6%. A further analysis showed that such a large drop can happen only when σ_v/σ_u is between 0.7 and 0.8. In such cases, the bias bottoms at $m = 5$ (if the true β is 2%) or at $m = 4$ (if the true β is 10% and σ_v/σ_u is 0.7), which is broadly consistent with the finding in Table 5. Assuming that σ_v/σ_u is 0.7, the non-skipping estimate (the case in which $m = 1$) is biased upwards by 14.11% (if the true β is 2%) or by 16.38% (if the true β is 10%). This can be seen from the first rows of Table 3C and 3D. Hence, in either case, it can be concluded that the non-skipping estimate is biased upwards by almost 15%.

Table 5: Evidence from the Japanese Prefectures

the estimated speed is in percentage

m	T_m	end year	$\hat{\beta}_m$	std.
1	39	90	16.88	1.18
2	38	89	18.65	1.34
3	39	90	12.00	1.00
4	36	87	13.69	1.13
5	35	86	12.84	1.14
6	36	87	10.27	1.02
7	35	86	11.02	1.09
8	32	83	12.59	1.27
9	36	87	12.56	1.29

Evidence from OECD countries

Data on GDP per capita for 24 OECD countries is taken from the Summers-Heston data set. The data is available for the period from 1950 to 1990. Hence, using the values for 1950 as initial values, in this case, $N = 24$ and $T = 40$. The result is summarized in Table 6. There is no clear tendency for the estimated speed to either go up or go down with m. In fact, given the relatively large standard errors, it could reasonably concluded that the estimate is more or less invariant to m. Hence, the result is not as clear cut as the previous two cases. But looking at Table 3C and 3D suggests that, as long as σ_v/σ_u is small, say less than 0.2, one should find a clear pattern of the estimated speed increasing with m. Hence, the result suggests that σ_v/σ_u should be at least larger than 0.2 even in this case. Even if it is 0.2, the non-skipping estimate would still be biased upwards by around 7% (the first rows in Table 3C and 3D). Hence, the conclusion is that the non-skipping estimate for OECD countries is likely to be biased upwards by at least around 7% even in this case.

Table 6: Evidence from OECD countries

the estimated speed is in percentage

m	T_m	end year	$\hat{\beta}_m$	std.
1	40	90	5.72	0.74
2	40	90	6.05	0.79
3	39	89	6.01	0.75
4	40	90	5.77	0.76
5	40	90	5.07	0.68
6	36	86	6.16	0.76
7	35	85	6.10	0.77
8	40	90	5.18	0.67
9	36	86	6.20	0.78
10	40	90	5.12	0.67

5 Conclusion

This paper has shown that the speed of convergence estimated by OLS with Fixed Effects is subject to a large upward bias, not only because of the small sample bias but also because of the measurement error bias.

Researchers who use the panel data approach to study convergence criticize the cross sectional approach for not using all the data available, and thus throwing away a lot of potentially important information that is there. They argue that the panel data approach is superior because it makes use of all the information in the data. However, the convergence equation is a growth theory equation. It is not supposed to explain year-to-year behaviors in output per capita. The measurement error bias, when measurement errors are defined in the broad sense, is a bias that comes from using too rich information for a simple model. Use of a rich data set would require a rich modelling of economies that goes beyond a simple convergence equation that can be derived from, for example, the Solow-Swan model.

Appendix

A Biases in the Skipping Estimation

To derive the bias in the estimated coefficient γ_m, the AR(1) coefficient in the skipping estimation, note first that equation 2 is no different from the usual OLS with fixed effects model except that the length of each period is m instead of 1. So, the formula for the biases in OLS with Fixed effect, which is presented in Section 2 of the text, can be directly applied. All one needs to do is to replace "T" in the original formula by "T/m". Also, $1-\beta$ has to be replaced by γ_m, and σ_u, the standard deviation of the true shock, has to be replaced by σ_{em}. Then one obtains

$$\hat{\gamma}_m - \gamma_m = -(B_{m1} + B_{m3})/(B_{m2} + B_{m4})$$

where

$$B_{m1} = \frac{1}{(T/m)^2} \cdot \frac{(T/m - 1) - (T/m) \cdot \gamma_m + {\gamma_m}^{T/m}}{(1-\gamma_m)^2}$$

$$B_2 = \frac{1}{1 - {\gamma_m}^2} \cdot \left\{ 1 - \frac{1}{T/m} - \frac{2 \cdot \gamma_m}{(1-\gamma_m)^2} \cdot \frac{(T/m - 1) - (T/m) \cdot \gamma_m + {\gamma_m}^{T/m}}{(T/m)^2} \right\}$$

$$B_{m3} = \left(\gamma_m - \gamma_m \cdot \frac{1}{T/m} + \frac{T/m - 1}{(T/m)^2} \right) \cdot \left(\frac{\sigma_v}{\sigma_{em}} \right)^2$$

and

$$B_{m4} = \left(1 + \frac{1}{T/m} \right) \cdot \left(\frac{\sigma_v}{\sigma_{em}} \right)^2.$$

In the above, σ_{em}^2 is the variance of e_{mit} which is

$$\sigma_{em}^2 = \frac{1 - (1-\beta)^{2m}}{1 - (1-\beta)^2} \cdot \sigma_u^2.$$

As I estimate the AR(1) coefficient by taking

$$\hat{\beta}_m = 1 - \hat{\gamma}_m^{1/m}$$

and noting that

$$\beta = 1 - \gamma_m^{1/m},$$

the bias in the estimated β is

$$\hat{\beta}_m - \beta = -\left[(1-\beta)^m - (B_{m1} + B_{m3})/(B_{m2} + B_{m4}) \right]^{1/m} + (1-\beta).$$

B Standard Errors for the Skipping Estimator

The original (non-skipping) model is

$$y_{it} = \rho \cdot y_{it-1} + \alpha_i + u_{it}$$

where ρ is defined as $1-\beta$. Define $\tilde{y}_i$ as the deviation of y_{it} from its regional (country) mean. Then the above model can be rewritten as

$$\tilde{y}_{it} = \rho \cdot \tilde{y}_{it-1} + u_{it}. \tag{3}$$

This model can be estimated by OLS, which is equivalent to maximum likelihood. The skipping model can be written as

$$\tilde{y}_{it} = \gamma \cdot \tilde{y}_{it-m} + e_{it} \tag{4}$$

where

$$\gamma = \rho^m \tag{5}$$

and

$$e_{it} = u_{it} + \rho \cdot u_{it-1} + \rho^2 \cdot u_{it-2} + \cdots + \rho^{m-1} \cdot u_{it-m+1}. \tag{6}$$

Equation (4) can be estimated by OLS to yield $\hat{\gamma}$ and $\hat{\sigma}_e^2$, and they are their respective maximum likelihood estimators. Note that maximizing the likelihood with respect to γ and σ_e^2 is equivalent to maximizing it with respect to ρ and σ_u^2. Hence, by using equations (5) and (6) on $\hat{\gamma}$ and $\hat{\sigma}_e^2$, we can derive the maximum likelihood estimators for ρ and σ_u^2 which I will denote $\hat{\rho}$ and $\hat{\sigma}_u^2$.

An asymptotic variance for $\hat{\rho}$ is given by

$$AVAR(\hat{\rho}) = -\left[\frac{\partial^2 \ln L}{\partial(\rho)^2}\right]^{-1}_{\rho=\hat{\rho},\sigma_e^2=\hat{\sigma}_e^2}$$

where $\ln L$ is the log likelihood function of the model (4). The question is what the derivative inside the bracket is. Applying the chain rule for the second derivative,

$$\begin{aligned}\frac{\partial^2 \ln L}{\partial(\rho)^2} &= \left[\frac{\partial^2 \ln L}{\partial(\gamma)^2} \cdot \left(\frac{\partial \gamma}{\partial \rho}\right)^2 + \frac{\partial \ln L}{\partial \gamma} \cdot \frac{\partial^2 \gamma}{\partial(\rho)^2}\right] \\ &+ \left[\frac{\partial^2 \ln L}{\partial\left(\sigma_e^2\right)^2} \cdot \left(\frac{\partial\left(\sigma_e^2\right)}{\partial \rho}\right)^2 + \frac{\partial \ln L}{\partial\left(\sigma_e^2\right)} \cdot \frac{\partial^2\left(\sigma_e^2\right)}{\partial(\rho)^2}\right]\end{aligned}$$

When evaluated at the maximum of the likelihood function, the second term inside each of the squared brackets is equal to zero, so the above expression simplifies into

$$\frac{\partial^2 \ln L}{\partial(\rho)^2} = \frac{\partial^2 \ln L}{\partial(\gamma)^2} \cdot \left(\frac{\partial\gamma}{\partial\rho}\right)^2 + \frac{\partial^2 \ln L}{\partial\left(\sigma_e^2\right)^2} \cdot \left(\frac{\partial\left(\sigma_e^2\right)}{\partial\rho}\right)^2. \tag{7}$$

I will first deal with derivatives of γ and σ_e^2 with respect to ρ. From equation (5), the first one is given by

$$\frac{\partial\gamma}{\partial\rho} = m \cdot \rho^{m-1}.$$

Next, from equation (6),

$$\sigma_e^2 = \left[1 + \rho + \rho^2 + \rho^4 + \cdots + \rho^{2\cdot(m-1)}\right] \cdot \sigma_u^2.$$

Hence,

$$\begin{aligned} \frac{\partial\left(\sigma_e^2\right)}{\partial\rho} &= \frac{1}{\rho} \cdot \left[\sum_{i=1}^{m-1} 2 \cdot i \cdot \rho^{2\cdot i}\right] \cdot \sigma_u^2 \\ &= \frac{1}{\rho} \cdot \left[\sum_{i=1}^{m-1} 2 \cdot i \cdot \rho^{2\cdot i}\right] \cdot \left[\sum_{j=1}^{m} \rho^{2(j-1)}\right]^{-1} \cdot \sigma_e^2. \end{aligned}$$

For the second derivatives of the log likelihood function with respect to γ and σ_u^2 in equation (7), evaluated at the maximum of the likelihood function, I approximate them by the negative of the inverse of their estimated variance:

$$-\left[\frac{\partial^2 \ln L}{\partial(\gamma)^2}\right]^{-1}_{\rho=\hat{\rho},\sigma\cdot_e^2=\hat{\sigma}\cdot_e^2} = AVAR\left(\hat{\gamma}\right) \approx V\hat{A}R\left(\hat{\gamma}\right)$$

and

$$-\left[\frac{\partial^2 \ln L}{\partial(\sigma_e^2)^2}\right]^{-1}_{\rho=\hat{\rho},\sigma\cdot_e^2=\hat{\sigma}\cdot_e^2} = AVAR\left(\hat{\sigma}_e^2\right) \approx V\hat{A}R\left(\hat{\sigma}_e^2\right) = \frac{2 \cdot \hat{\sigma}_e^4}{N \cdot T_m}.$$

Hence,

$$\begin{aligned} AVAR\left(\hat{\rho}\right)^{-1} &= \left[m \cdot \hat{\rho}^{m-1}\right]^2 \cdot V\hat{A}R\left(\hat{\gamma}\right)^{-1} \\ &\quad + \frac{N \cdot T_m}{2} \cdot \left\{\frac{1}{\hat{\rho}} \cdot \left[\sum_{i=1}^{m-1} 2 \cdot i \cdot \hat{\rho}^{2\cdot i}\right] \cdot \left[\sum_{j=1}^{m} \hat{\rho}^{2(j-1)}\right]^{-1}\right\}^2. \end{aligned}$$

C Data Sources

US data

The original data is all from Bureau of Economic Analysis (BEA). For 1929-1992, I downloaded the data from Xavier Sala-i-Martin's home page. I combined the two series that appear in his file using the growth rate. For 1992-1996, I downloaded the data from the home page of BEA. I linked this series with the previous series using the growth rate.

Japanese Data

The original data is almost entirely from the Economic Planning Agency (EPA) of Japan, *Prefectural Economic Accounts*. However, between 1950 and 1955, values for some prefectures are missing from the official data. Miyohei Shinohara et. al., *Chiiki Kozo no Keiryo Bunseki* (Econometric Analysis of Regional Structure, in Japanese) estimates these missing values. I use these estimated numbers whenever the official numbers were not available. I linked slightly different series using the growth rates at: 1965, 1975 and 1980. Note on the data for 1950: The 1950 value for Tokyo, the wealthiest of the prefectures, appeared to be problematic. In 1950, its total prefectural income, according to the data, was 565.5 (in billions of yen). In 1951, it went down to 513.2, while the rest of Japan grew at the rate of 26.3% (nominal). Between 1951 and 1952 no such obvious anomaly was found. Hence, I concluded that the data for 1950 was distorted, perhaps due to a typing mistake, and decided not to use it. Hence, the whole analysis for Japan starts in 1951.

OECD data

All the data is taken from the Summers and Heston data set, available from the home page of the University of Toronto. I used the series coded RGDPL, real GDP per capita. I excluded member countries that participated OECD recently, for the sake of comparison with other research and also for the sake of data availability. The 24 countries in my sample are: Canada, USA, Austria, Belgium, Denmark, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, UK, Japan, Finland, Australia, and New Zealand.

References

Anderson, T.W. and C. Hsiao, "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, 76, 598-606, 1981.

Arellano, Manuel, and Stephen Bond, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58, 277-297, 1991.

Barro, Robert J. and Xavier Sala-i-Martin, "Convergence", *Journal of Political Economy*, vol. 100, 223-251, April 1992.

Barro, Robert J. and Xavier Sala-i-Martin, *Economic Growth*, McGraw Hill, 1995.

Canova, Fabio and Albert Marcet, "The Poor Stay Poor: Non-Convergence across Countries and Regions", Universitat Pompeu Fabra working Paper 137, October 1995.

Caselli, Francesco, Gerardo Esquivel and Fernando Lefort, "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics", *Journal of Economic Growth*, 1, 363-389, September 1996.

De la Fuente, Angel, "On the Sources of Convergence: A Close Look at the Spanish Regions", mimeo, Universitat Autònoma de Barcelona, October 1996.

Evans, Paul, "How to Estimate Growth Regressions Consistently", mimeo, Ohio State University, January 1995.

Hsiao, Cheng, *Analysis of Panel Data*, Cambridge University Press, Cambridge, UK, 1986.

Islam, Nazrul, "Growth Empirics: A Panel Data Approach", *Quarterly Journal of Economics*, 110, 1127-1170, 1995.

Knight, Malcolm, Norman Loayza and Delano Villanueva, "Testing the Neoclassical Theory of Economic Growth: A Panel Data Approach", *IMF Staff Papers*, 40, 512-541, September 1993.

Nerlove, Marc, "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections", *Econometrica*, 39, no. 2, 359-382, March 1971.

Nickell, Stephen, "Biases in Dynamic Models with Fixed Effects", *Econometrica*, 49, no. 6, 1417-1426, November, 1981.

Solow, Robert M., "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics*, 70, 65-94, February 1956.

Swan, Trevor W., "Economic Growth and Capital Accumulation", *Economic Record*, 32, 334-361, November 1956.