Using panel methods to detect convergence in economic growth

Thijs Knaap*

August 23rd, 1996

Abstract

The Solow model of economic growth is briefly explained, as is the principle of *general methods of moments* estimation. Methods are then derived to estimate and test the shrinking parameter in a simple model of growth. The data are a panel of 56 countries with each 124 annual observations, but with a fair percentage missing observations. The methods of estimation are adapted to this incomplete panel by splitting incomplete time series into smaller ones. Weighted *GMM* is compared to *unweigted least squares*. We find considerable small sample bias in the former method. Finally, the hypothesis of a unit root cannot be rejected.

* The author wishes to thank Tom Wansbeek for fruitful comments. Any remaining errors are mine.

1 Growth: economic theory and econometric practice

In this section, we will first take a brief look at the theory of economic growth. Several theories offer hypotheses about the behavior of a country's output in time. These hypotheses have all been confronted with economic data before. In the second part of this section, we will look at the different ways in which hypotheses have been tested.

1.1 The theory of economic growth

A lot has changed since 1896. After a century of continuous changes in the way people live, work and think, we find ourselves in a situation that is very much different from the last *fin de siècle*. Undeniably, progress has taken place in a lot of different ways. One of the most visible changes is a large increase in production. If we were to compare the amount of goods and services each Dutch person, on average, produces every year we would find that this amount has grown at a rate of approximately 1.7 % since 1896. Add to that an average population growth of 1.2 % per year, and we find that our total annual production is more than 15 times that of a century ago.

The theory of economic growth tries to identify the sources of this increase in production. At the root, economic growth comes from an increase in production capacity. This capacity is determined by a number of constraints. First, increments in the size of the population and its productive component, the labor force, are a continuous source of growth. Secondly, the amount of capital goods accumulated accounts for an important part of the capacity to produce. Thirdly, the state of technology determines the productivity of labor and capital. Technological progress can enhance our abilities to produce goods and services. As a fourth component of the production capacity, the economic institutional environment is often included. Political stability and well-defined rights and duties stimulate prosperity. A fifth constraint for economic growth is the amount of human capital present; this incorporates the fact that an uneducated labor force is not usually as productive as a group of educated workers.

1.1.1 The Solow model

The implications of the different sources of growth can be readily understood with the help of a rather simple mathematical model. The following was first published by Robert Solow in 1956, but it remains at the centre of much of today's research. We will assume a production function for a closed economy that incorporates the country's capital stock K_t, its labor force L_t and its level of technology A_t, all at time t.

$$Y_t = F(A_t, L_t, K_t) \tag{1}$$

Production can be used for two purposes, consumption and capital accumulation. The proportion that is used for the latter is denoted by s, the saving rate. It is assumed fixed and exogenous. Capital depreciation is modelled as the annual disappearance of a fraction δ of the entire stock.

$$\frac{\mathrm{d}K_t}{\mathrm{d}t} = sY_t - \delta K_t \tag{2}$$

The rate of population growth and the rate of growth of A_t, the level of technology, are taken fixed and exogenous at n and x, respectively. The production function $F(\cdot)$ is neoclassical: it exhibits positive and diminishing marginal products with respect to K and L, has constant returns to scale and satisfies the *Inada conditions*[1].

1.1.2 Steady state and dynamic behavior

In general, technological growth is taken to be of the *labor augmenting* form. This means that $F(A_t, L_t, K_t)$ can be written as $F(A_t \cdot L_t, K_t)$. Now introduce the following notation: for a variable Y, the per capita value Y/L is written as y. The value of Y per effective unit of labor, $Y/(A \cdot L)$, is denoted $\hat{y}$. Because of the assumed constant returns to scale, we can rewrite (1) as

$$\hat{y}_t = F(\hat{k}_t, 1) \equiv f(\hat{k}_t) \tag{3}$$

Using (2) and the chain rule for differentiating, we find that

$$\frac{\mathrm{d}\hat{k}_t}{\mathrm{d}t} = sf(\hat{k}_t) - (\delta + x + n)\hat{k}_t \tag{4}$$

Because of the regularity conditions we imposed on $F(\cdot)$, there are two values of $\hat{k}$ for which the expression in (4) is zero: $\hat{k} = 0$ and another positive value $\hat{k}^*$. It is easy to see that there must hold $\frac{\mathrm{d}f}{\mathrm{d}\hat{k}}(\hat{k}^*) < 1$. Therefore, the following must be true:

$$\begin{aligned} \frac{\mathrm{d}\hat{k}_t}{\mathrm{d}t} &> 0 \quad \text{if} \quad \hat{k}_t < \hat{k}^* \\ \frac{\mathrm{d}\hat{k}_t}{\mathrm{d}t} &< 0 \quad \text{if} \quad \hat{k}_t > \hat{k}^* \end{aligned}$$

and we can see that, regardless of the value of $\hat{k}_0$, the economy will move toward a *steady state* at which $\hat{k}_t = \hat{k}^*$.

We can see three sources of economic growth at work here. First, if we assume that in the initial state of affairs $\hat{k}_0 < \hat{k}^*$, we see an increase in $\hat{k}_t$ through time. By relation (3), we see that this means an increase in $\hat{y}_t$. Second, in the *steady state*, $\hat{y}_t$ is constant. This means, however, that per capita production y_t is growing at rate x because of technological improvements. Thirdly, population growth contributes to the fact that total production Y_t is growing at rate $x + n$ in the *steady state*.

1.1.3 Convergence

If we now compare two closed economies i and j and assume, following Solow, that the coefficients x, n, δ and s are the same in both countries, we can deduce that both economies will end up in the same *steady state* in which $\hat{k}_t = \hat{k}^*$. This will happen regardless of the initial values $\hat{k}_{i,0}$ and $\hat{k}_{j,0}$.

If we want to quantify the rate of convergence, that is, how quickly the economies are approaching their *steady state*, we need to specify the production function in formula (1). Supposing we use a Cobb-Douglas form, we would get:

$$Y_t = A_t L_t^{1-\alpha} K_t^{\alpha} \tag{5}$$

[1] These conditions are: $\lim_{K\to 0}(F_K) = \lim_{L\to 0}(F_L) = \infty$ and $\lim_{K\to\infty}(F_K) = \lim_{L\to\infty}(F_L) = 0$

Defining $\gamma_{\hat{k}_t}$ as the rate of growth of $\hat{k}_t$, that is, $\frac{\mathrm{d}\hat{k}_t/\mathrm{dt}}{\hat{k}_t}$, we can use equation (4) to write

$$\gamma_{\hat{k}_t} = sA_t^{\alpha}\hat{k}_t^{-(1-\alpha)} - (x+n+\delta) \tag{6}$$

If we now assume that $\hat{k}_t$ is close enough to k^* to approximate equation (5) with a first order Taylor series after rewriting it as a function of log $(\hat{k}_t)$, we find that

$$\gamma_{\hat{k}_t} \approx -(1-\alpha)(x+n+\delta)\log\frac{\hat{k}_t}{\hat{k}^*} \tag{7}$$

The coefficient $(1-\alpha)(x+n+\delta)$ is usually termed β in growth economics. It is a measure of the speed of convergence; the bigger β, the faster an economy will move toward the *steady state*.

1.2 Testing the convergence hypothesis

The strong result of the preceding paragraph, that the per capita level of the GDP of different countries converges in time, can be qualified to some extent. Namely, if the values of the parameters δ, n and s vary over countries, then so will the values of $\hat{k}^*$. A weaker version of the convergence hypothesis called *conditional convergence* is defined as the situation in which GDP per capita, conditional on the values of δ, n and s, shows convergence. Both the absolute and conditional convergence hypothesis have been tested extensively and in a number of ways. In general, three different hypotheses are distinguished:

1. *β-convergence* is defined as a situation in which poor countries in general grow faster than rich countries, so that there is a genuine 'catch-up' process. Tests of this hypothesis are known as 'Barro-regressions'
2. *σ-convergence* means that the variance of the variable GDP per capita in a sample of countries decreases over time.
3. *Unit root tests* search for non-stationarity in (possibly detrended) time series of GDP per capita. These tests can be done for whole panels as well as single countries. Finding a unit root is seen as an indication that no convergence is taking place.

The different tests are discussed in the following sub-paragraphs.

1.2.1 Barro-regressions

Economists Mankiw, Romer and Weil (*MRW*) started their 1992 article with a much quoted remark about 'taking Solow seriously', in face of growing scepticism about the model. It is one of the best known articles in which so-called 'Barro-regressions' are done. These regressions take the form

$$\ln\left(\frac{y_t}{y_0}\right) = X\theta + \lambda\ln(y_0) + \epsilon_{it} \tag{8}$$

with X containing variables that explain the level of the *steady state*. It is taken as evidence in support of the convergence hypothesis if λ is significantly negative, as this would indicate that countries with a low GDP per capita generally grow faster than richer countries.

The findings of *MRW* and most other authors using these techniques can be summarized as follows:

- If unconditional convergence is tested (leaving $X\theta$ out of the equation) λ is only significantly negative in small, coherent subsets of countries like the OECD. For big samples, like all the countries in the world, the hypothesis $\lambda = 0$ cannot be rejected. When testing convergence conditional on a number of country-specific regressors, a significantly negative λ is found.

Conditional on each country's savings, population growth and stock of human capital (proxied by the number of people in secondary school) the estimated λ is -0.289 [0.062], implying a β of 0.0137. The authors take this as proof of conditional convergence. Within the OECD, $\hat{\beta}$ is as much as 0.0203 [0.002]. The results indicating a negative value of λ are generally known as results implying *β-convergence*. It is the view of a great many authors that *β-convergence* is a real world phenomenon. This does not neccesarily imply that *σ-convergence* is, though. Mathematically, it can be shown that *β-convergence* is only a neccesary, not a sufficient condition.

1.2.2 σ-convergence

If we take a look at our 56-country dataset, we see that the variance of the logarithm of GDP per capita increases by half over the period 1951-1992. This means that whatever is going on in this panel, it is not σ-convergence. Economist Danny Quah (1993) asserts that taking *β-convergence* as evidence for *σ- convergence* is the classic 'Galton's fallacy of regression towards the mean'. To underline that process of divergence is actually taking place, he describes the process of changing GDP per capita's as a Markov chain. The 'states' in the chain are determined by a country's y divided by the world average GDP per capita: $[0, \frac{1}{4}), [\frac{1}{4}, \frac{1}{2}), [\frac{1}{2}, 1), [1, 2)$ and $[2, \infty)$. The matrix of estimated transition probabilities for a period of 23 years is

$$\begin{pmatrix} 0.76 & 0.52 & 0.09 & 0 & 0 \\ 0.12 & 0.31 & 0.20 & 0 & 0 \\ 0.12 & 0.10 & 0.46 & 0.24 & 0 \\ 0 & 0.07 & 0.26 & 0.53 & 0.05 \\ 0 & 0 & 0 & 0.24 & 0.95 \end{pmatrix} \tag{9}$$

with the equilibrium distribution over the states being

$$(0.16, 0.05, 0.10, 0.12, 0.57) \tag{10}$$

Quah concludes that there may be significant *β-convergence*, but the sample of countries in the world can still be modelled as one that drifts apart in two categories: poor and rich.

1.2.3 Panel data approaches

In this sub-paragraph we will briefly discuss two papers that explicitly use panel methods to measure convergence. They differ in the available data and the approach taken to estimation.

Knight, Loayza and Villanueva, 1993

The authors (*KLV*) operate with a balanced panel of 98 countries and 5 observations each 5 years apart. Their data starts in 1960. The model *KLV* use is much like the one in paragraph 1.1.1, only they include human capital as an explicit factor of production and allow the state of technology A_t to be partly determined by the degree of openness to foreign trade (F) and the level of government fixed investment (P).

For estimation, the authors employ the Chamberlain Π method. This works as follows: first, the vector of observations for each country (z_i) is regressed on all the explanatory variables x_i using the SUR method:

$$\begin{pmatrix} z_1 \\ \vdots \\ z_T \end{pmatrix} = \Pi \begin{pmatrix} x_1 \\ \vdots \\ x_T \end{pmatrix}$$

Here, the x_t are vectors with containing each period's regressors. It turns out that Π is a 13×6 matrix. Using the model, the structure of Π is restricted and the implied model parameters are estimated by GMM:

$$\min_{\Psi} \left(\Pi - f(\Psi)\right)' \Omega^{-1} \left(\Pi - f(\Psi)\right)$$

with Ω the variance of Π obtained by using SUR. The advantage over the cross-sectional approach of *MRW* is that the existence of country-specific effects, possibly correlated with explanatory variables, is taken into full account. Their results include an estimated β of 0.05 [t=8.32], which is a lot higher than *MRW*'s. This, according to the authors, is due to the very correlation between the individual effects and y_{i0} we just mentioned. The coefficients of other explanatory variables each have the expected sign.

Lee, Pesaran and Smith, 1995

In this working paper the authors *(LPS)* implement ideas about estimation using dynamic panel data that were developed in a previous article (Pesaran and Smith, 1995). The following important conclusions are reached:

- Regressions of the type (8) are very badly biased in their estimation of γ and are not informative about convergence;
- The economic model can be written in the following reduced form:

$$\begin{aligned} \ln(y_{it}) &= \delta_i + x_i t + u_{it} \\ \delta_i &= \frac{\mu_i - x_i \gamma_i}{1 - \gamma_i} \\ u_{it} &= \gamma_i u_{i,t-1} + \epsilon_{it} \end{aligned}$$

  The standard Solow-model allows γ_i to vary across countries and constrains g_i to be equal across the world. *LPS* estimate the model by maximum likelihood and examine the effects of restricting either of these variables to be the same across countries. Table 1 shows the results: if an estimate is allowed to vary across countries, the average value is shown. The number in brackets is the standard error. The sample is that of 102 non-oil countries over 30 years.
  It seems that heterogenous panel bias can have quite an impact on estimations. Restricting one parameter to be zero severely affects the other estimated parameters.

- The cross-country variance does not decrease over time; the hypothesis of no σ-convergence cannot be rejected.
- The so-called *t-bar test* from Im, Pesaran and Shin (1995) is used to test whether the set of the individual country's γ_i is significantly smaller than unity on average. The unit root hypothesis could not be rejected. This is an example of the third kind of convergence testing, mentioned in the beginning of this paragraph.

Table 1: LPS estimation results

Restricted parameter	$\hat{\gamma}$	$\hat{g}$
None	0.7600 [0.0124]	0.01737 [0.0005]
g	0.9032 [0.0114]	0.02025 [0.0011]
γ	0.8315 [0.0144]	0.01732 [0.0006]
γ, g	0.9589 [0.0050]	0.02179 [0.0006]

2 Estimation

In this section, we will derive the method by which our model shall be estimated. We concentrate on the *minimum distance* estimator or the Generalized Method of Moments (GMM). This because it allows maximum freedom in specifying the model while delivering asymptotically efficient results. The philisophy behind this method is introduced in the following section. We will thereafter concentrate on the applications of GMM on dynamic panel data.

In section 2.2 we use the working hypothesis that a full panel is available. In practice, this will not be the case; measures to deal with missing observations are discussed in section 2.3.

2.1 Minimum distance estimation

The distinguishing characteristic of panel data as a separate category is that it runs in two dimensions. Assume, for example, a sampling process in which observations are numbered $1, \ldots, N$ (the first dimension) and in which an observation looks like

$$r_i' = [X_{i,1}, X_{i,2}, \ldots, X_{i,T}] \tag{11}$$

Assume r_i is i.i.d. from some multivariate distribution with finite fourth moments[2], and further suppose observations $X_{i,t}$ are actually vectors consisting of the dependent variables $y_{i,t}$ and regressors $x_{i,t}$. Define w_i as the vector of elements of $r_i \otimes r_i$ with nonzero variance, let $\mu = \mathrm{E}w_i$ and let $\bar{w}$ be the average over the w_i's. If we assert there exists a linear relation between the dependent variables and the regressors, it makes sense to consider the minimum mean-square error linear predictors of $y_{i,t}$,

$$\mathrm{E}^*(y_{i,t}|x_{i,t}) = \pi' x_{i,t}$$

It is clear that π is a function of μ, since

$$\pi' = \mathrm{E}(y_{i,t}x_{i,t}') \left[\mathrm{E}(x_{i,t}x_{i,t}')\right]^{-1}$$

Practical estimates of π are obtained using the sample equivalents of these expectations which form, of course, the familiar least squares estimator:

$$\hat{\pi} = \left[\left(\sum_{i=1}^{N} x_{i,t}x_{i,t}'\right)^{-1} \sum_{i=1}^{N} x_{i,t}y_{i,t}\right]$$

[2]This derivation is based on Chamberlain (1984) and Wansbeek and Kapteyn (1994)

So far, not very much seems to be gained with this approach. The advantages become clear however when we want to restrict the model's second order parameters. This is equivalent to restricting the structure of μ by demanding it is a function of a lower-dimensional vector θ. This way, model-specific covariance structures can be imposed on the estimation procedure[3].

Estimation now amounts to finding θ^0, the true value of θ. We therefore employ the minimum distance estimator which chooses $\hat{\theta}$ such that

$$\min_{\hat{\theta}} \left[\bar{w} - f(\hat{\theta})\right]' A_n \left[\bar{w} - f(\hat{\theta})\right] \tag{12}$$

with A_n symmetric, $A_n \to \Psi$ almost surely and the asymptotic weighting matrix Ψ positive definite.

Because $\bar{w}$ is the average of a series of i.i.d. observations, we can invoke the Kolmogorov[4] law of large numbers II to assert $\bar{w} \to \mu$ almost surely as $N \to \infty$. As $\mu = f(\theta^0)$ we can see that the expression in (12) converges in probability to

$$\left[f(\theta^0) - f(\hat{\theta})\right]' \Psi \left[f(\theta^0) - f(\hat{\theta})\right] \tag{13}$$

which is minimized for only one value in a neighborhood of θ^0, and that is θ^0 itself. This implies the consistency of $\hat{\theta}_{mde}$. As for this estimator's asymptotic distribution, in section 2.5 it is shown that $\sqrt{N}\left(\hat{\theta}_{mde} - \theta^0\right) \to \mathrm{N}(0, \Lambda)$, with

$$\Lambda = (F_0'\Psi F_0)^{-1} F_0'\Psi \mathrm{V}(w_i)\Psi F_0 (F_0'\Psi F_0)^{-1} \tag{14}$$

with $\mathrm{V}(w_i)$ the variance matrix of w_i. Because this matrix is positive definite, there holds that $\Lambda - (F'(\mathrm{V}(w_i))^{-1}F)^{-1}$ is positive semi-definite. Therefore, an optimal choice for Ψ is $\mathrm{V}^{-1}(w_i)$. This is the case we shall refer to when we speak of the *minimum distance estimator.* If $A_n = I$, the unit matrix, the estimator is known as *Unweighted Least Squares (ULS).*

Small sample problems with GMM

Most of the GMM characteristics are derived in a limiting case. As $N \to \infty$, the method is unbiased and has the distribution in (14). However, Altonji and Segal (1995) argue that estimating $\mathrm{V}(w_i)^{-1}$ with the same data that is used in minimizing the expression in (12) may cause small sample biases. He presents an intuitive argument that runs as follows: suppose we try to estimate a variance θ with observations D_{pi}, with $i = 1, \ldots, N$ and $p = 1, \ldots, N_i$. The observed variances are

$$m_i = \frac{1}{N_i - 1} \sum_{p=1}^{N_i} (D_{pi} - \bar{D}_i)^2$$

[3] About the function f we assume continuity and at least double differentiability around θ^0. We denote

$$F = F(\theta) \equiv \frac{\mathrm{d}f(\theta)}{\mathrm{d}\theta'}, \hat{F} \equiv F(\hat{\theta}), F_0 \equiv F(\theta^0)$$

$F(\theta)$ is assumed to be of full column rank for θ in a neighborhood of θ^0.

[4] For this and the Lindeberg-Lévy C.L.T., see Amemiya, 1985

and they are independent but heteroscedastic. The variances of the $m_i's$ are estimated as

$$\hat{\omega}_i = \frac{N_i^2}{(N_i-1)(N_i-2)^2}\left[\frac{1}{N_i}\sum_{p=1}^{N_i}(D_{pi}-\bar{D}_p)^4 - \left(\frac{1}{N_i}\sum_{p=1}^{N_i}(D_{pi}-\bar{D}_p)^2\right)^2\right]$$

He argues the $\hat{\omega}_i$'s are positively correlated with the m_i's in small samples because outliers in the D_{pi} tend to increase both expressions. The bias in $\hat{\theta}$ can then be written as

$$\hat{\theta} - \theta = \left(\sum_{i=1}^{N}\omega_i^{-1}\right)^{-1}\sum_{i=1}^{N}\omega_i^{-1}(m_i - \mathrm{E}m_i) \tag{15}$$

The first term is always positive, and in the second the smaller values of m_i receive a bigger weight than the larger ones. They therefore tend to cause a downward bias.

The positive correlation 'intuitively' introduced by Altonji can of course be checked using our data. We will return to this subject in the next section.

2.2 Dynamic models and estimation

We will consider two models to describe the behavior of a country's GDP per capita through time; the first model ignores the linear trend growth due to technological improvements:

$$y_{it} = \gamma y_{i,t-1} + \lambda + \alpha_i + u_{it} \tag{16}$$

The second model incorporates the exogenous growth as well, thereby introducing the exogenous variable 'time':

$$y_{it} = \gamma y_{i,t-1} + \lambda + g \cdot t + \alpha_i + u_{it} \tag{17}$$

In both equations, y_{it} is the natural logarithm of country i's GDP per capita in year t. Two error terms are involved in the equations, the individual effect α_i and regular error term u_{it}. The errors are generated as

$$\alpha_i \sim \mathrm{N}(0, \sigma_\alpha^2)$$
$$u_{it} \sim \mathrm{N}(0, \sigma_u^2)$$

Assume for a minute that both σ_u^2 and σ_α^2 are zero. This effectively eliminates both error terms from the model. In model (16), the parameter γ determines the existence of $\lim_{t\to\infty} y_{it}$. It is assumed γ is positive; if $\gamma \geq 1$, the limit does not exist. If $\gamma = 1$, y_{it} behaves linearly in time, its exact path depending on the value of λ. Finally, if $0 < \gamma < 1$, the limit is equal to $\lambda/(1-\gamma)$. In the latter case, we can talk about convergence in per capita income. In model (17) we also allow for exogenous technological growth if parameter $g > 0$. Then, an economy's steady state amounts to that situation where growth of per capita income is linear with rate $g/(1-\gamma)$. This state is attained if $0 < \gamma < 1$.

We shall first deal with the estimation of model (16). Then, in section 2.4 we shall extend the model as in equation (17) and enhance our method of estimation. For now, however, we will work with the first model only.

Call the number of countries in our panel N, each with T annual observations. Define the vectors y_i as the observations for country i, y_{it}, stacked from $t = 1, \ldots, T$. Now construct the following $(T-1) \times T$ matrix

$$D \equiv \begin{pmatrix} -\gamma & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & -\gamma & 1 \end{pmatrix} = I_R - \gamma I_L \tag{18}$$

implicitly defining matrices I_R and I_L. Create $(T-1) \times 1$ vectors ϵ_i as Dy_i.

Each element ϵ_{it} is equal to $\lambda + \alpha_i + u_{it}$ and thus has expectation λ. We will first discuss a method of estimation that removes the term $\lambda + \alpha_i$ from ϵ_{it}. In the next sextion, we will attempt to estimate λ and σ^2_α separately.

Introduce the $(T-1) \times (T-1)$ 'within'-matrix W,

$$W \equiv I_{T-1} - \frac{1}{T-1}\left(\iota_{T-1}\iota'_{T-1}\right), \qquad \iota = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

This matrix 'removes the average' of a vector. Defining $\epsilon^*_i = W\epsilon_i$, we see that $\mathrm{E}\epsilon^*_i = 0$ and $\mathrm{E}(\epsilon^*_i)(\epsilon^*_i)' = W'\Sigma_u W = \sigma^2_u W$, with $\Sigma_u = \sigma^2_u \cdot I$. Therefore, the variance of ϵ^*_i is a function of σ^2_u. We can write this function as

$$\mathrm{E}\ \mathrm{vec}\left[(\epsilon^*_i)(\epsilon^*_i)'\right] = \mathrm{E}\epsilon^*_i \otimes \epsilon^*_i = \sigma^2_u \mathrm{vec}\left[W\right] R\sigma^2_u$$

with $R = \mathrm{vec}(W)$.

With this specification we arrive at the situation that we started with in equation (11). We have N countries that each supply a (supposedly) i.i.d. time series, whose variance after an algebraical transformation conforms to a theoretical structure. All we need to do now is find the value of σ^2_u that minimizes a criterion like (12). Because this criterion happens to be linear in σ^2_u, that can be done by projecting the transformed data on the linear subspace orthogonal to the one spanned by the columns of R. Let M_R be the projection matrix orthogonal to R taking redundant restrictions due to symmetry into account:

$$M_R = \frac{1}{2}(I_{(T-1)^2} + P_{T-1,T-1}) - R(R'R)^{-1}R'$$

with $P_{T-1,T-1}$ the symmetric commutation matrix of order $(T-1)^2 \times (T-1)^2$. Define variables m_i as

$$m_i \equiv M_R(\epsilon^*_i \otimes \epsilon^*_i)$$

Our model then dictates that

$$\mathrm{E}[\frac{1}{N}\sum_{i=1}^{N} M_R(\epsilon^*_i \otimes \epsilon^*_i)] = \mathrm{E}\frac{1}{N}\sum_{i=1}^{N} m_i = 0 \tag{19}$$

This expression can be seen in the light of equation (12) as follows: the 'unity part' of M_R supplies a term resembling $\bar{w}$, whereas the 'projection part' fits the optimal θ, optimal in the least squares sense. We still have one parameter left unspecified, however; when using the matrix D to compute ϵ^*_i, we implicitly used the parameter γ. In order to find this parameter, we have to directly invoke the GMM principle and minize the expression

$$\bar{m}' A_n \bar{m} \tag{20}$$

with $\bar{m} \equiv \frac{1}{N}\sum_{i=1}^{N} m_i$ and A_n an estimator of the generalized inverse of the variance of m_i. The former expression can be written as

$$\bar{m} = M_R(W \otimes W)(D \otimes D)\frac{1}{N}\sum_{i=1}^{N}(y_i \otimes y_i) \tag{21}$$

The latter expression involves the variance of m_i. That variance is formally defined as $\mathrm{E}((m_i - \mathrm{E}m_i)'(m_i - \mathrm{E}m_i))$. However, since $\mathrm{E}m_i = 0$, we can write

$$\begin{aligned}\mathrm{V}(m_i) = \mathrm{E}(m_i'm_i) &= \mathrm{E}\left(P(Dy_i \otimes Dy_i)(Dy_i \otimes Dy_i)'P'\right)\\ &= \mathrm{E}\left(P(Dy_iy_i'D' \otimes Dy_iy_i'D')P'\right)\end{aligned}$$

denoting $P \equiv M_R(W \otimes W)$ for notational convenience. The sample counterpart of this expectation constitutes our estimator of $\mathrm{V}(m_i)$:

$$\widehat{\mathrm{V}(m_i)} = \frac{1}{N}\sum_{i=1}^{N}\left(P(\hat{D}y_iy_i'\hat{D}' \otimes \hat{D}y_iy_i'\hat{D}')P'\right) \tag{22}$$

Notice that the dimensions of this matrix in the context of long macro-economic time series are huge. With our T well in excess of 100, it looks like computational problems might drive us in the arms of ULS rather than GMM.

Finally, the standard error of this estimator, given in formula (14), can be simplified somewhat depending on our choice for Ψ, the asymptotic weighting matrix. Take a look at the expression in (14) to see that we can write

$$\sqrt{N}(\hat{\gamma}_{mde} - \gamma) \to \mathrm{N}\left(0, (\kappa_\gamma'\mathrm{V}^+(m_i)\kappa_\gamma)^{-1}\right)$$

with $\kappa_\gamma \equiv \mathrm{E}(\partial m_i/\partial\gamma)$ if we use what we have called the *GMM*-estimator. If we use *unweigthed least squares* we have

$$\sqrt{N}(\hat{\gamma}_{mde} - \gamma) \to \mathrm{N}\left(0, \frac{\kappa_\gamma'\mathrm{V}(m_i)\kappa_\gamma}{(\kappa_\gamma'\kappa_\gamma)^2}\right)$$

In the following, again define $P \equiv M_R(W \otimes W)$ for convenience.

$$\begin{aligned}m_i &= P(Dy_i) \otimes (Dy_i)\\ &= P((I_R - \gamma I_L)y_i) \otimes ((I_R - \gamma I_L)y_i)\\ &= P\left[(I_Ry_i \otimes I_Ry_i - \gamma I_Ly_i \otimes I_Ry_i - \gamma I_Ry_i \otimes I_Ly_i + \gamma^2 I_Ly_i \otimes I_Ly_i\right]\end{aligned}$$

and thus

$$\begin{aligned}\frac{\partial m_i}{\partial\gamma} &= P\left[-I_Ly_i \otimes I_Ry_i - I_Ry_i \otimes I_Ly_i + 2\gamma I_Ly_i \otimes I_Ly_i\right]\\ &= -P\left[I_Ly_i \otimes [I_R - \gamma I_L]y_i + [I_R - \gamma I_L]y_i \otimes I_Ly_i\right]\\ &= -P\left[I_Ly_i \otimes Dy_i + Dy_i \otimes I_Ly_i\right]\end{aligned}$$

the last step using the equality $D = I_R - \gamma I_L$ from equation (18). Using the same equality, we can continue

$$\begin{aligned}\frac{\partial m_i}{\partial\gamma} &= -\frac{1}{\gamma}P\left[-I_Ry_i \otimes Dy_i - Dy_i \otimes Dy_i + Dy_i \otimes I_Ry_i - Dy_i \otimes Dy_i\right]\\ \mathrm{E}\frac{\partial m_i}{\partial\gamma} &= -\mathrm{E}\frac{1}{\gamma}P\left[I_Ry_i \otimes Dy_i + Dy_i \otimes I_Ry_i\right]\end{aligned}$$

the last step using the restrictions implied in equation (19). If we use the fact that $\bar{m}$ is equal to the average of the m_i's and insert the estimated value of γ, we can actually estimate the derivative:

$$\begin{aligned} \hat{\kappa}'_\gamma &= -\frac{1}{\hat{\gamma}} P \frac{1}{N} \sum_{i=1}^{N} \left[I_R y_i \otimes \hat{D} y_i + \hat{D} y_i \otimes I_R y_i \right] \\ &= -\frac{1}{\hat{\gamma}} P \left[I_R \otimes \hat{D} + \hat{D} \otimes I_R \right] \left[\frac{1}{N} \sum_{i=1}^{N} y_i \otimes y_i \right] \end{aligned} \quad (23)$$

This way, we have a ready-to-use expression for the standard error of our estimate of γ, without having to bother with numerical derivatives.

2.3 Missing observations

In the previous section, it was assumed that a balanced panel of observations was available. In the perilous world of international macro-economic reality, however, we often find panels with over half the observations unavailable. In our panel, just under 35 % of the data is not there, rendering the methods of the last section useless unless adapted.

The adaptations to cope with missing data are introduced in this section. We will actually show *two* ways in which the researcher can tackle the problem. As it turns out, one of these methods clearly outperforms the other in our specific context.

Method # 1: Pretending to have a full panel

Look again at formula (21). For estimation purposes, all the available data is condensed in the variable $\bar{m}$. If we could find a way to create a missing-observations version of this variable, we can proceed with the rest of the approach as if nothing were the matter. And, as a matter of fact, we can do just that by simply taking the average over the values that are there, and neglecting the values that aren't.

Define variables $k_{i,j,m}$ as follows:

$$k_{i,j,m} = \begin{cases} 1 & \text{if both } y_{i,j} \text{ and } y_{i,m} \text{ are there.} \\ 0 & \text{if } y_{i,j} \text{ or } y_{i,m} \text{ is not there.} \end{cases}$$

with $i = 1, \ldots, N$ and $j, m = 1, \ldots, T-1$. Now define $K_{j,m} \equiv \sum_{i=1}^{N} k_{i,j,m}$ and K the $(T-1)^2 \times 1$ vector of $K_{j,m}$'s. Substitute 0 for all missing observations and we have our missing-variables version of (21):

$$\bar{m}^* = M_R (W \otimes W)(D \otimes D) \left[\text{diag}(K) \right]^{-1} \sum_{i=1}^{N} (y_i \otimes y_i) \quad (24)$$

The same variable could be used in equations (22) and (23).

However, the story is not over yet. When using patchup-methods like this, we have to remember the implicit assumptions that are being made. In this case, the core assumption is that the average over the available observations instead of over all the individuals has the same expectation as the latter. This because the accuracy of our estimator derives directly from this expectation. This assumption could be violated if the absence of observations were somehow correlated with explanatory variables of our model. Suppose for instance that we tried to explain consumer expenditure on food by the consumer's income, and that poor people categorically refused to reveal their income. This would leave us with a panel of just rich people, whose influence of income on food expenditure is, by Engel's law, far less. Biased coefficient estimates may thus result from these latent correlations.

The influence of correlated missing observations on dynamic panel data are even more intricate. It is hard to give a general direction of the bias caused by this problem. Our feeling when using the above method was that it underestimated the parameter γ somewhat. The following serves to illustrate how such underestimation may come about.

Looking at the pattern of our data, it seems that less developed countries have more observations missing than western industrial panel members. The observations that are available for those less developed countries are clustered at the end and at the beginning of our period of observation. Both these characteristics make sense. With development comes a better administration of a country's production and population, which explains the tendency for richer countries to be able to supply more data. Also, as other countries develop through time, their ability to supply data increases which explains the clustering at the end. The clustering at the beginning is a feature of the specific research report we used, which tried to give some idea of world development over the years and estimated GDP per capita for *all* countries for a number of key years, some of them in the first observational round.

All in all, we may conclude that there is reason to believe that the values of y_{it} may be positively correlated with the probability that country i's GDP per capita is observed in year t. Looking at the values taken on by the average of the available observations each year, it seems that after an initial low value, there is a period in which only the western nations supply data and the average sores. This period is then followed by a decline because of increasing availability of other country's data.

We now formalize this pattern in an, admittedly, crude way. Suppose we only have data from one individual for three time periods. Call the observations y_1, y_2 and y_3. We now estimate the parameter γ imposing the model $y_t = \gamma y_{t-1} + \lambda + u_t$ by the above method. This means we calculate γ as

$$\min_{\gamma} \bar{m}'\bar{m} \tag{25}$$

$$\bar{m} = M_R \left[W \begin{bmatrix} -\gamma & 1 & 0 \\ 0 & -\gamma & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \right] \otimes \left[W \begin{bmatrix} -\gamma & 1 & 0 \\ 0 & -\gamma & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \right]$$

In this case,

$$W = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}, \qquad M_R = \begin{bmatrix} \frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

We can elaborate $\bar{m}$ as

$$\frac{1}{2} M_R \left[(\gamma(y_2 - y_1) + (y_2 - y_3)) \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right] \otimes \left[(\gamma(y_2 - y_1) + (y_2 - y_3)) \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right]$$

$$= \frac{1}{2} M_R(\gamma^2(y_2 - y_1)^2 - 2\gamma(y_3 - y_2)(y_2 - y_1) + (y_3 - y_2)^2) \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \tag{26}$$

so that minimizing (25) is the same as minimizing $f(\gamma)$ with

$$\begin{aligned} f(\gamma) &= \gamma^2(y_2 - y_1)^2 - 2\gamma(y_2 - y_1)(y_3 - y_2) + (y_3 - y_2)^2 \\ \frac{\mathrm{d}f(\gamma)}{\mathrm{d}\gamma} &= 2\gamma(y_2 - y_1) - 2(y_2 - y_1)(y_3 - y_2) \end{aligned} \tag{27}$$

This means we can write $\hat{\gamma}$ as $(y_3 - y_2)/(y_2 - y_1)$. Remember now that our data showed an unnatural increase in the middle period. In this stylized example, that would mean an exorbitant value for y_2, which has the effect of lowering the estimate of γ and introducing a negative bias.

Method # 2: Tiny time series

If you use a time series with annual observations that spans more than a century you are bound to run into some missing observations. Large parts of the series may be intact, though. It is this phenomenon, combined with the fact that the model does not change over time, that we use in the second method.

Modify the panel in the following way: every time series y_i is split up into several smaller parts, all with the same length T^*, with $T^* \geq 3$. Some data may be lost in the process, e.g. if the number of observations does not divide by T^* we lose some border observations, or if there exist sub-series smaller than T^*.

We now have a complete panel of observations, with a larger N than before and a smaller T, equal to T^*. We need not worry about any problems this method may cause in terms of bias: the same model applies to the small series as to the larger ones. There are some possible drawbacks, though:

- The possible loss of data if T^* is an awkward number. This is, of course, in the hands of the researcher.
- The artificial nature of the panel may reflect upon the estimates of the parameters' standard errors. This because the variation present in the panel is now partly created by the researcher himself. The direction of this effect is not entirely clear from the outset.
- Another point of weakness may be the possibility of domination of the data by countries that have a relatively large amount of data available. If it should be so that countries which are poor have relatively large gaps in their time series, this may influence our results. This is especially so if those poor countries inherently also have a different value of γ, instead of the assumed uniformity of this parameter.

Of course, this method of estimation has some advantages as well: apart from having a complete panel with short time series, we also experience an increase in the number of ‘individuals’[5]. This may improve our estimates as they were derived asymptotically as $N \to \infty$.

2.4 Practical estimation and Exogenous Variables

With either of the two methods for dealing with missing observations, we find ourselves having to optimize a criterion like (20), with A_n the unit matrix. This section deals with the practical aspects of that optimization, as well as with the incorporation of exogenous variables in that routine. We shall first assume a situation in which $\bar{m}$ is given by equation 21, and no exogenous variables are present. Then we consider the case where exogenous variables are included in our model. All the other cases can then easily be derived.

So, firstly, our criterion looks like $\min_\gamma \bar{m}'\bar{m}$, with

$$\bar{m} = M_R(W \otimes W)(D \otimes D)S \tag{28}$$

with S summarizing our data. If we want to write $\bar{m}$ as a function of γ, we split up the expression into

$$\bar{m} = M_R(W \otimes W)\left(\gamma^2(I_L \otimes I_L)S - \gamma(I_L \otimes I_R)S - \gamma(I_R \otimes I_L)S + (I_R \otimes I_R)S\right)$$

For ease of notation, we rewrite this into an expression with three new variables S_0, S_1, S_2 that each contain some of the data:

$$\bar{m} = \gamma^2 S_2 + \gamma S_1 + S_0$$

The definitions of the S_x are implied. The form at which we have now arrived enables us to write (28) explicitly as a function of γ:

$$\min_\gamma S_0'S_0 + \gamma \cdot 2S_1'S_0 + \gamma^2 \cdot (S_1'S_1 + 2S_2'S_0) + \gamma^3 \cdot 2S_1'S_2 + \gamma^4 \cdot S_2'S_2$$

With all the S_x known, this criterion is easily minimized using an analytical first and second derivative in the Gauss-Newton rootfinding procedure.

We now turn to our second case, in which dispose of the premultiplying matrix W and incorporate exogenous variables x_i. We now need to minimize a different criterion:

$$\min_{\gamma,g} \bar{m}_x'\bar{m}_x \tag{29}$$

$$\bar{m}_x \equiv \frac{1}{N}\sum_{i=1}^{N} M_R\left[(Dy_i - gx_i) \otimes (Dy_i - gx_i)\right]$$

$$D \equiv \begin{pmatrix} -\gamma & 1 & & \\ & \ddots & \ddots & \\ & & -\gamma & 1 \end{pmatrix} \equiv -\gamma I_L + I_R$$

This criterion can again be split up into a polynomial in the variables g and γ, summarizing the data in six new variables $T_1, \ldots, T_6$.

$$\bar{m} = T_1 + gT_2 + \gamma T_3 + g\gamma T_4 + g^2 T_5 + \gamma^2 T_6$$

[5]This term becomes more and more abstract as we continue our modification of the data. We mean the number of observed time series in the panel.

$$
\begin{aligned}
T_1 &= M_R(I_R \otimes I_R)\sum_{i=1}^{N}(y_i \otimes y_i) \\
T_2 &= -M_R(I_R \otimes I_R)\sum_{i=1}^{N}(y_i \otimes x_i + x_i \otimes y_i) \\
T_3 &= -M_R(I_R \otimes I_L + I_L \otimes I_R)\sum_{i=1}^{N}(y_i \otimes y_i) \\
T_4 &= M_R[(I_R \otimes I_L)\sum_{i=1}^{N}(x_i \otimes y_i) + (I_L \otimes I_R)\sum_{i=1}^{N}(y_i \otimes x_i)] \\
T_5 &= M_R(I_R \otimes I_R)\sum_{i=1}^{N}(x_i \otimes x_i) \\
T_6 &= M_R(I_L \otimes I_L)\sum_{i=1}^{N}(y_i \otimes y_i)
\end{aligned}
$$

All the T_x can readily be computed provided the data are complete. They also allow us to analytically write out criterion (29)

$$
\begin{aligned}
\bar{m}'\bar{m} &= a_1 + a_2 g + a_3 g^2 + a_4 g^3 + a_5 g^4 + a_6\gamma + a_7\gamma^2 + a_8\gamma^3 + \\
&+ a_9\gamma^4 + a_{10}\gamma g + a_{11}\gamma^2 g + a_{12}\gamma^3 g + a_{13}\gamma g^2 + a_{14}\gamma^2 g^2 + a_{15}\gamma g^3
\end{aligned}
$$

$$
\begin{array}{lcllcl}
a_1 &=& T_1'T_1 & a_9 &=& T_6'T_6 \\
a_2 &=& 2T_2'T_1 & a_{10} &=& 2T_1'T_4 + 2T_2'T_3 \\
a_3 &=& T_2'T_2 2T_5'T_1 & a_{11} &=& 2T_2'T_6 + 2T_3'T_4 \\
a_4 &=& 2T_2'T_5 & a_{12} &=& 2T_6'T_4 \\
a_5 &=& T_5'T_5 & a_{13} &=& 2T_5'T_3 + 2T_2'T_4 \\
a_6 &=& 2T_1'T_3 & a_{14} &=& T_4'T_4 \\
a_7 &=& T_3'T_3 + 2T_1'T_6 & a_{15} &=& 2T_4'T_5 \\
a_8 &=& 2T_3'T_6 & & &
\end{array}
$$

Again, using theoretical partial first and second derivatives, this criterion can readily be minimized with the two-dimensional Gauss-Newton procedure.

2.5 Appendix

Because $\bar{w}$ is the average of a series of i.i.d. observations, we can use the Lindeberg-Lévy C.L.T. to claim

$$\sqrt{N}(\bar{w} - \mu) \to \mathrm{N}(0, \mathrm{V}(w_i)) \tag{30}$$

with $\mathrm{V}(w_i)$ the variance matrix of w_i.

For the following, it is important to recognize that $\hat{\theta}_{mde}$ really is a function of $\bar{w}$. We will therefore explicitly introduce the function $\theta(x) : \Re^q \to \Upsilon \subset \Re^p$, which is defined by

$$\min_{\hat{\theta}\in\Upsilon} \Delta(\theta, \bar{w}) = \min_{\hat{\theta}\in\Upsilon} \left[\bar{w} - f(\hat{\theta})\right]' A_n \left[\bar{w} - f(\hat{\theta})\right] \tag{31}$$

Note that $\theta(f(\theta^0)) = \theta^0$. We can use relation (30) and the δ-method to assert

$$\sqrt{N}\left(\hat{\theta}_{mde} - \theta^0\right) \to \mathrm{N}\left(0, \left.\frac{\mathrm{d}\theta(x)}{\mathrm{d}x'}\right|_{x=\mu} \mathrm{V}(w_i) \left.\frac{\mathrm{d}\theta'(x)}{\mathrm{d}x}\right|_{x=\mu}\right) \tag{32}$$

In order to find an expression for the derivatives in this formula's right hand side we use the first-order condition implied by the minimization in (31) to write

$$\frac{\partial \Delta(\theta(\cdot), \bar{w})}{\partial \theta(\cdot)} = -2 \cdot \hat{F}' A_n (\bar{w} - f(\hat{\theta})) \equiv r(\hat{\theta}, \bar{w}) = 0$$

Differentiating both sides of this equation to $\bar{w}'$, we get

$$\frac{\partial^2 \Delta(\theta(\cdot), \bar{w})}{\partial \theta(\cdot) \partial \theta'(\cdot)} \cdot \frac{\partial \theta(\bar{w})}{\partial \bar{w}'} + \frac{\partial^2 \Delta(\theta(\cdot), \bar{w})}{\partial \theta(\cdot) \partial \bar{w}'} = 0$$

or, in the notation with $r(\hat{\theta}, \bar{w})$,

$$\frac{\partial r(\theta(\cdot), \bar{w})}{\partial \theta(\cdot)'} \cdot \frac{\partial \theta(\bar{w})}{\partial \bar{w}'} + \frac{\partial r(\theta(\cdot), \bar{w})}{\partial \bar{w}'} = 0$$

This can then rewritten to

$$\frac{\partial \theta(\bar{w})}{\partial \bar{w}'} = -\left(\frac{\partial r(\theta(\cdot), \bar{w})}{\partial \theta(\cdot)'} \right)^{-1} \frac{\partial r(\theta(\cdot), \bar{w})}{\partial \bar{w}'} \tag{33}$$

This form implies that any constants in front of $r(\cdot, \cdot)$ will cancel; we therefore strike the -2 factor for now. The two matrices on the right can be computed as follows: we see that

$$\begin{aligned} \frac{\partial r(\theta, \cdot)}{\partial \theta'} = &= -\hat{F}' A_n \hat{F} + \frac{\partial}{\partial \theta'} \left(\hat{F}' A_n (\bar{w} - f(\cdot)) \right) \\ &= -\hat{F}' A_n \hat{F} + \frac{\partial}{\partial \theta'} \left(\text{vec} \left[(\bar{w} - f(\cdot))' A_n \cdot \hat{F} \cdot I \right] \right) \\ &= -\hat{F}' A_n \hat{F} + (I \otimes (\bar{w} - f(\cdot) A_n)) \frac{\partial \text{vec} \hat{F}}{\partial \theta'} \end{aligned}$$

so $\text{plim} \frac{\partial r(\theta, \cdot)}{\partial \theta'} = -F_0' \Psi F_0$. The second matrix in (33) can also be computed:

$$\frac{\partial r(\cdot, x)}{\partial x'} = \hat{F}' A_n$$

straightforwardly, so $\text{plim} \frac{\partial r(\cdot, x)}{\partial x'} = -F_0' \Psi$. Inserting these two values implies

$$\text{plim} \frac{\partial \theta(\bar{w})}{\partial \bar{w}'} = (F_0' \Psi F_0)^{-1} F_0' \Psi$$

This now allows us to explicitly write out the asymptotic distribution of $\hat{\gamma}_{mde}$.

3 Results

In this section we will showcase the results from implementing the various methods of the last section on our data. First, the data is introduced in the next section. Then, we estimate the two models from the last section using the two methods for coping with missing observations. We conclude in section 3.3

3.1 The data

The data comes from Maddison (1995), and forms a panel of 56 countries for which the GDP per capita is monitored in the years 1820, 1850 and 1870 thru 1993. In our research, we omit 1820 and 1850, thus only using each country's last 124 years. Of the 6944 possible observations, 4564 are actually there. Figure **??** shows each year's average y and the number of observations available for that year. Notice that the gaps in our data occur especially in the beginning of the sample period. There are some spikes in the number of observations, where y was estimated for almost the whole panel in some key year. Whenever these spikes occur, we see a drop in the average y, indicating that in the beginning of the sample period, the poorer countries tend to supply less data than the richer countries.

For some of the methods of estimation, we are required to convert the panel into one with smaller time series and a larger number of individuals. Table 2 shows how the number of observations and complete time series varies with the different length of those 'tiny time series'

Table 2: Data availability

Series length	# of series	Total # of observations
3	1467	4401
4	1087	4348
5	856	4280
6	713	4278
7	606	4242
8	525	4200
9	461	4149
10	421	4210
20	201	4020
30	125	3750
40	97	3880
50	45	2250

As one would expect, the number of observations used decreases with the length of the time series. Requiring long series also has the effect of eliminating a lot of data from developing countries, who are often unable to supply lengthy series.

3.2 Estimation results

3.2.1 Method # 1: Pretending to have a full panel

Using this first method, described in section 2.3, with no exogenous technological growth, gave the following results:

$$\begin{aligned} \hat{\gamma} &= 0.9193 \\ \sqrt{\mathrm{V}(\hat{\gamma})} &= 2.1942 \cdot 10^{-3} \end{aligned}$$

As we already noted, this result may be biased downwards somewhat.

3.2.2 Method #2: Tiny Time Series

We estimated the model with and without exogenous technological growth, with *ULS* and *GMM* and with a number of different sizes of the 'tiny time series'. The results are summarized in the following tables.

Table 3: Exogenous Growth, ULS

Series length	$\hat{\gamma}$	[standard error]	$\hat{g}$	[standard error]
3	1.0865	[...]	-0.0074221	$[6.7895 \cdot 10^{-6}]$
4	1.0301	$[1.8631 \cdot 10^{-7}]$	-0.0019693	$[2.6375 \cdot 10^{-6}]$
5	0.9869	$[1.9907 \cdot 10^{-7}]$	0.0017753	$[1.2882 \cdot 10^{-6}]$
6	1.0056	$[3.3805 \cdot 10^{-7}]$	-0.0001370	$[6.9192 \cdot 10^{-7}]$
7	0.9897	$[4.5296 \cdot 10^{-7}]$	0.0008094	$[6.5062 \cdot 10^{-7}]$
8	0.9979	$[1.6397 \cdot 10^{-7}]$	0.0011053	$[5.5438 \cdot 10^{-7}]$
9	1.0047	$[1.2424 \cdot 10^{-7}]$	-0.0003490	$[2.9144 \cdot 10^{-7}]$
10	1.0012	$[1.1976 \cdot 10^{-7}]$	-0.0000898	$[1.8252 \cdot 10^{-7}]$
11	0.9983	$[2.5726 \cdot 10^{-7}]$	0.0004098	$[2.5937 \cdot 10^{-7}]$

Table 4: Exogenous Growth, GMM

Series length	$\hat{\gamma}$	[standard error]	$\hat{g}$	[standard error]
3	1.0865	[...]	-0.007422	[...]
4	0.6127	[0.14681]	0.041776	[0.008894]
5	0.8487	[0.00156]	0.010930	$[9.9713 \cdot 10^{-5}]$
6	1.0737	[0.00113]	-0.009877	$[7.7316 \cdot 10^{-5}]$
7	1.1711	[0.00159]	-0.018259	$[9.6677 \cdot 10^{-5}]$
8	0.8855	[0.00026]	0.005904	$[1.4874 \cdot 10^{-5}]$
9	1.2151	[0.00059]	-0.023064	$[3.6425 \cdot 10^{-5}]$
10	1.3729	[0.00044]	-0.006792	$[2.0655 \cdot 10^{-5}]$
11	0.8572	[0.00020]	0.007480	$[9.0109 \cdot 10^{-6}]$

Tables 3 thru 6 allow for a few careful conclusions. First of all, it appears that for these small values of N, *ULS* works better than *GMM*. The estimates from the former fluctuate less over the different lengths of the time series. However, the small standard errors seem a little exaggerated when looking at the different parameter estimates in the same table. Secondly, we have a case of heterogeneous panel bias: leaving out one parameter (exogenous growth) clearly influences the remaining estimates. Thirdly, the question whether γ is significantly smaller than unity cannot be answered from these tables. They either show $\hat{\gamma}$ both significantly above *and* below one in the same table, or are inconclusive. To test for the presence of a unit root in our data, we introduce a method to combine the observations.

Table 5: No Exogenous Growth, ULS

Series length	$\hat{\gamma}$	[standard error]
3	0.7514	[0.0661]
4	0.9990	[0.0117]
5	1.0030	[0.0044]
6	1.0043	[0.0025]
7	0.9969	[0.0029]
8	1.0072	[0.0040]
9	1.0016	[0.0012]
10	1.0004	[0.0006]
11	1.0016	[0.0011]

Table 6: No Exogenous Growth, GMM

Series length	$\hat{\gamma}$	[standard error]
3	0.7514	[0.0686]
4	1.1183	[0.0162]
5	0.9021	[0.0138]
6	0.8482	[0.0208]
7	0.8580	[0.0136]
8	1.0326	[0.0070]
9	1.0130	[0.0019]
10	1.1533	[0.0135]
11	0.9481	[0.0071]

3.2.3 Fisher's P_λ unit root test

R.A. Fisher (1932) introduces a method to combine several independent unit root tests into one single number. This test uses the observed significance level of the different tests, P_i. This 'p-value' is the probability that the result occured under the hypothesis of a unit root. Under this hypothesis and random sampling the 'p-value' has a uniform distribution on $[0, 1]$. Combining the different P_i as $\chi^2 = -2 \sum \ln(P_i)$ gives a statistic that, under the null hypothesis, has an exact χ^2 distribution with $2N$ degrees of freedom.

We used the method of tiny time series described above to compute $\hat{\gamma}$ and its standard error for all 56 countries separately. We then compute each country's P value under the hypothesis of a unit root. These P values are combined into a statistic χ^2_t. For different sizes of the time series, the different results are in table 7. In this table, estimation is done using *ULS*. Using *GMM* on these small samples resulted in such things as negative estimates of the standard error.

It seems that the hypothesis of a unit root is rejected if we use really small time series (size 3 and 4). For the other lengths, the null of a unit root cannot be rejected. This still leaves our question about the possibility of $\gamma = 1$ unanswered, though. However, that can be remedied by just using *all* the P-values from the different countries and time series sizes and computing

Table 7: Unit root tests

Series length (t)	χ^2_t	P value using $\chi^2(112)$
3	221.853	$3 \cdot 10^{-9}$
4	151.089	0.0082
5	77.949	0.9940
6	106.933	0.6176
7	113.085	0.4535
8	109.039	0.5616
9	121.403	0.2560
10	83.662	0.9791
11	81.483	1.0000

one statistic. Under the null of a unit root, this statistic has a $\chi^2(1008)$ distribution. Using all the data created to construct the previous table, we computed $\chi^2 = 1046.497$, so that the 'p-value' under the null is around 0.194. Clearly, this is not small enough to reject the null hypothesis of a unit root.

3.3 Conclusions

The methods derived and applied in the previous sections all in all appear to be a bit volatile. The estimate obtained from method # 1 is conspicuously low, and those from method # 2 suffer from heterogeneous panel bias and vary significantly over the different time series lengths. However, Fisher's P_λ unit root test allows us to combine the different pieces of information into a single test statistic so that we may provide an answer to the question 'to convergence or not to converge'. With the available data and methods, the answer is clear: we are not able to reject the hypothesis 'not to converge'.

4 Literature

Altonji, J.G. and L.M. Segal [1995]: 'Small sample bias in GMM estimation of covariance structures', unpublished paper, Northwestern University, August 1995 draft.

Amemiya, T. [1985]: *Advanced Econometrics*, Harvard University Press, Cambridge MA

Barro, R.J. and X. Sala-i-Martin [1992]: 'Convergence', *Journal of Political Economy*, 100, pp. 223 - 251

Barro, R.J. and X. Sala-i-Martin [1995]: *Economic Growth*, McGraw-Hill, New York

Bernard, A.B. and S.N. Durlauf [1996]: 'Interpreting tests of the convergence hypothesis', *Journal of Econometrics*, 71, pp. 161 - 173

Chamberlain, G. [1984]: 'Panel Data', in Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics, Volume II*, North Holland, Amsterdam, pp. 1248 - 1318

Fisher, R.A. [1932]: *'Statistical Methods for Research Workers'*, Oliver and Boyd, Edinburgh 4th ed.

Harvey, A.C. [1993]: *Time Series Models*, Harvester Wheatsheaf, Hertfordshire U.K.

Im, K., M.H. Pesaran and Y. Shin [1995]: 'Testing for unit roots in Dynamic heterogenous Panels', *DAE Working Paper*, University of Cambridge, U.K.

Knight, M., N. Loayza and D. Villanueva [1993]: 'Testing the Neoclassical Theory of Economic Growth', *IMF Staff Papers*, 40, pp. 512 - 541

Lee, K., M.H. Pesaran and R. Smith [1995]: 'Growth and convergence: A multi-country empirical analysis of the Solow growth model', *DAE Working Paper 9531*, University of Cambridge, U.K.

Maddala, G.S. and P.C. Liu [1996]: 'Panel data unit root tests: What do they test?', unpublished paper, Ohio State University, March 1996 draft.

Maddison, A. [1995]: *Monitoring the World Economy, 1820-1992*, UITGEVER, PLAATS

Malinvaud, E. [1970]: *Statistical Methods of Econometrics*, North-Holland, Amsterdam

Mankiw, N.G., D. Romer and D.N. Weil [1992]: 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics*, 107, pp. 407 - 437

Pesaran, M.H. and R. Smith [1995]: 'Estimating long-run relationships from dynamic heterogeneous panels', *Journal of Econometrics*, 68, pp. 79 - 113

Quah, D. [1993]: 'Galton's fallacy and Tests of the Convergence Hypothesis', *Scandinavian Journal of Economics*, 95, pp. 427 - 443

Sevestre, P. and A. Trognon [1992]: 'Linear Dynamic Models', in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data*, Kluwer Academic Publishers, Dordrecht, pp. 95 - 117

Solow, R.M. [1956]: 'A Contribution to the Theory of Economic Growth', *Quarterly Journal of Economics*, 70, pp. 65 - 94

Wansbeek, T. and A. Kapteyn [1994]: *Measurement Error and Latent Variables in Econometrics*, unpublished.