

Models for Converging Economies

Andrew Harvey and Vasco Carvalho
Faculty of Economics, University of Cambridge

March 1, 2002

Abstract

The aim of this article is the development of models for converging economies. After discussing models of balanced growth, univariate models of the gap between per capita income in two economies are examined. The preferred models combine unobserved components with an error correction mechanism and allow a decomposition into trend, cycle and convergence components. A new type of second-order error correction mechanism is shown to be particularly useful in this respect. The levels of per capita income in two economies may be modelled jointly by bivariate convergence models. These models generalise balanced growth models and can be based on autoregressive or unobserved components formulations. Both approaches provide coherent forecasts but the unobserved components models also yield a description of trends, cycles and convergence components. The methods are applied to data on the US and Japan. The generalisation to multivariate series is then set out.

KEYWORDS: Cycles; balanced growth; error correction mechanism; stochastic trend; unobserved components.

1. Introduction

The aim of this paper is to suggest models for converging economies, set out estimation procedures and apply them to real data. The preferred models combine unobserved components with an error correction mechanism and allow a decomposition into trend, cycle and convergence components. This provides insight into what has happened in the past and gives a coherent procedure for the prediction of future observations.

The starting point is the literature on convergence. Barro and Sala-i-Martin (1992) adopted a cross-sectional approach to determine whether the growth rate of a country depended on the gap between its income per capita and that of the US. Subsequently many researchers fitted autoregressive time series models containing a mechanism in which growth effectively depends on the gap. However, this mechanism was rarely the focus of attention. Instead, the concern was with tests of convergence ; see, for example, Bernard and Durlauf (1996) and Evans and Karras (1996).

The way in which structural, or unobserved components (UC), time series models, can be used to model trends and cycles in multivariate series is set out in section 2. Issues of stability are then addressed. An understanding of stability is essential to modelling convergence and the balanced growth models described in section 2 lay the foundation for much of what is to come later. Section 3 examines definitions of convergence in the literature and suggests some modifications. Simply fitting a smooth trend to the difference between two economies is proposed as a way of establishing stylised facts about convergence. Because the trend is based on a model, some direct tests on the extent to which convergence has actually taken place are possible. A dynamic error correction model for convergence is then proposed and extended so as to incorporate a mechanism which allows convergence to take place smoothly. These models are fitted to data on GDP in the US and Japan. The information gained from fitting autoregressive models is also assessed.

The levels of per capita income in two economies may be modelled jointly by bivariate convergence models. These models, introduced in section 4, generalise balanced growth models and can be based on autoregressive or unobserved components formulations. The models for the difference of two economies, set out in section 3, can be derived from the bivariate models. The properties of the bivariate models are explored and they are fitted to the Japanese and US series. Section 5 generalises the ideas of section 4 to more than two series. The properties of these multivariate models are set out and the implications of various special cases are explored.

A subsidiary aim of the paper is to clarify the issues involved in testing for convergence and stability. The questions: Are we in a stable state? and Are we converging to a stable state? are completely different. It is argued that stationarity tests are more appropriate than unit root tests when the null hypothesis is that the series are stable. The unit root tests for convergence are best understood in the context of a model for the dynamics of convergence of two series. However,

the fitting of models of convergence allows the researcher to address more interesting questions on the speed of convergence and the size of the gap between two economies which have converged.

2. Structural Time Series Models and Balanced Growth

2.1. Univariate models

The *local linear trend* model for a set of observations, $y_t, t = 1, \dots, T$, consists of stochastic trend and irregular components, that is

$$y_t = \mu_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (2.1)$$

The trend, μ_t , receives shocks to both its level and slope so

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim NID(0, \sigma_\eta^2), \\ \beta_t &= \beta_{t-1} + \zeta_t, & \zeta_t &\sim NID(0, \sigma_\zeta^2), \end{aligned} \quad (2.2)$$

where the irregular, level and slope disturbances, ε_t, η_t and ζ_t , respectively, are mutually independent and the notation $NID(0, \sigma^2)$ denotes normally and independently distributed with mean zero and variance σ^2 . If both variances σ_η^2 and σ_ζ^2 are zero, the trend is deterministic. When only σ_ζ^2 is zero, the slope is fixed and the trend reduces to a random walk with drift

$$\mu_t = \mu_{t-1} + \beta + \eta_t. \quad (2.3)$$

Allowing σ_ζ^2 to be positive, but setting σ_η^2 to zero gives an integrated random walk trend, which when estimated tends to be relatively smooth. The model is often referred to as the '*smooth trend*' model.

The statistical treatment of unobserved component models is based on the state space form (SSF). Once a model has been put in SSF, the Kalman filter yields estimators of the components based on current and past observations. Signal extraction refers to estimation of components based on all the information in the sample. Signal extraction is based on smoothing recursions which run backwards from the last observation. Predictions are made by extending the Kalman filter forward. Root mean square errors (RMSEs) can be computed for all estimators and prediction or confidence intervals constructed.

The unknown variance parameters are estimated by constructing a likelihood function from the one-step ahead prediction errors, or innovations, produced by the Kalman filter. The likelihood function is maximized by an iterative procedure; see Harvey (1989). The calculations can be done with the STAMP package of Koopman et al (2000). Once estimated, the fit of the model can be checked using standard time series diagnostics such as tests for residual serial correlation.

Distinguishing a long-term trend and from short-term movements is important. Short-term movements may be captured by adding a serially correlated stationary component, ψ_t , to the model. Thus

$$y_t = \mu_t + \psi_t + \varepsilon_t, \quad t = 1, \dots, T \quad (2.4)$$

An autoregressive process is often used for ψ_t . Another possibility is the stochastic cycle

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \quad t = 1, \dots, T, \quad (2.5)$$

where λ_c is frequency in radians and κ_t and κ_t^* are two mutually independent white noise disturbances with zero means and common variance σ_κ^2 . Given the initial conditions that the vector $(\psi_0, \psi_0^*)'$ has zero mean and covariance matrix $\sigma_\psi^2 \mathbf{I}$, it can be shown that for $0 \leq \rho < 1$, the process ψ_t is stationary and indeterministic with zero mean, variance $\sigma_\psi^2 = \sigma_\kappa^2 / (1 - \rho^2)$ and autocorrelation function

$$\rho(\tau) = \rho^\tau \cos \lambda_c \tau, \quad \tau = 0, 1, 2, \dots \quad (2.6)$$

For $0 < \lambda_c < \pi$, the spectrum of ψ_t displays a peak, centered around λ_c , which becomes sharper as ρ moves closer to one; see Harvey (1989, p60). The period corresponding to λ_c is $2\pi/\lambda_c$. In the limiting cases when $\lambda_c = 0$ or π , ψ_t collapses to first-order autoregressive processes with coefficients ρ and minus ρ respectively. More generally the reduced form is an *ARMA*(2, 1) process in which the autoregressive part has complex roots. The complex root restriction can be very helpful in fitting a model, particularly if there is reason to include more than one cycle.

Imposing the smooth trend restriction often allows a clearer separation into trend and cycle.

2.2. Multivariate models

Suppose we have N time series. Define the vector $\mathbf{y}_t = (y_{1t}, \dots, y_{Nt})'$ and similarly for $\boldsymbol{\mu}_t, \boldsymbol{\psi}_t$ and $\boldsymbol{\varepsilon}_t$. Then a multivariate UC model may be set up as

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\psi}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim NID(\mathbf{0}, \Sigma_\varepsilon), \quad t = 1, \dots, T, \quad (2.7)$$

where Σ_ε is an $N \times N$ positive sem-definite matrix. The trend is

$$\begin{aligned} \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim NID(\mathbf{0}, \Sigma_\eta) \\ \boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\zeta}_t, & \boldsymbol{\zeta}_t &\sim NID(\mathbf{0}, \Sigma_\zeta), \end{aligned}$$

With $\Sigma_\eta = 0$, we get the smooth trend model. With $\Sigma_\zeta = 0$, we get the random walk plus drift.

The *similar cycle* model, introduced by Harvey and Koopman (1997) is

$$\begin{bmatrix} \boldsymbol{\psi}_t \\ \boldsymbol{\psi}_t^* \end{bmatrix} = \begin{bmatrix} \rho \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \otimes \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{\psi}_{t-1} \\ \boldsymbol{\psi}_{t-1}^* \end{bmatrix} + \begin{bmatrix} \boldsymbol{\kappa}_t \\ \boldsymbol{\kappa}_t^* \end{bmatrix}, \quad t = 1, \dots, T, \quad (2.8)$$

where $\boldsymbol{\psi}_t$ and $\boldsymbol{\psi}_t^*$ are $N \times 1$ vectors and $\boldsymbol{\kappa}_t$ and $\boldsymbol{\kappa}_t^*$ are $N \times 1$ vectors of the disturbances such that

$$E(\boldsymbol{\kappa}_t \boldsymbol{\kappa}_t') = E(\boldsymbol{\kappa}_t^* \boldsymbol{\kappa}_t^{*'}) = \Sigma_\kappa, \quad E(\boldsymbol{\kappa}_t \boldsymbol{\kappa}_t^{*'}) = \mathbf{0}, \quad (2.9)$$

where Σ_κ is an $N \times N$ covariance matrix. The model allows the disturbances to be correlated across the series. Because the damping factor and the frequency, ρ and λ_c , are the same in all series, the cycles in the different series have similar properties; in particular their movements are centred around the same period. This seems eminently reasonable if the cyclical movements all arise from a similar source such as an underlying business cycle. Furthermore, the restriction means that it is often easier to separate out trend and cycle movements when several series are jointly estimated.

2.3. Example: US and Japan

Models with smooth trends were fitted to the logarithms of quarterly, seasonally adjusted, data on real GDP per capita in the US and Japan over the period 1961:1 to 2000:1. The data were obtained from the OECD Main Economic Indicators and

the population series was constructed from annual data by repeating each annual observation four times and then applying a four quarter moving average. The series are in 1990 US dollars; the choice of conversion date affects the gap between the series, but is otherwise irrelevant.

The cycle fitted to Japan is rather weak in the univariate model. By contrast, it becomes much more like the US cycle in the similar cycle bivariate model. Table 1 shows the estimates of the parameters, together with the standard error (SE) for each equation and the Box-Ljung statistic, $Q(P)$, based on the first P residual autocorrelations. The correlations between the slope, cycle and irregular disturbances were -0.143, 0.274 and 1 respectively. The period of 27.07 quarters corresponds to 6.77 years.

	<i>TABLE 1</i>	Bivariate	Model
	<i>Hyperparameters</i>	<i>Japan</i>	<i>US</i>
<i>Trend</i>	$\sigma_\zeta(\times 10^{-3})$	1.638	0.907
<i>Cycle</i>	$\sigma_\kappa(\times 10^{-3})$	7.177	7.642
	$\sigma_\psi(\times 10^{-3})$	17.22	18.34
	ρ_c	0.91	0.91
	<i>Period</i> ($2\pi/\lambda_c$)	27.07	27.07
<i>Irregular</i>	$\sigma_\varepsilon(\times 10^{-3})$	4.380	0.174
<i>Fit</i>	$\log L$	982.343	
	$SE(\times 10^{-3})$	11.144	9.058
<i>Diagnostics</i>	$Q(11)$	11.766	14.719

The cycles shown in figure 2.1 are for the bivariate model. Their presence means that the trends are quite smooth. However, it is clear that the forecasts will diverge as there is virtually no growth in Japan at the end of the series.

2.4. Stability and balanced growth

The *balanced growth* UC model is a special case of (2.7):

$$\mathbf{y}_t = \mathbf{i}\mu_t + \boldsymbol{\alpha} + \boldsymbol{\psi}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \quad (2.10)$$

where μ_t is a univariate local linear trend, \mathbf{i} is a vector of ones, and $\boldsymbol{\alpha}$ is an $N \times 1$ vector of constants. If μ_t is initialised with a diffuse prior, then $\boldsymbol{\alpha}$ must be subject to a constraint so it contains only $N - 1$ free parameters, for example there may be one zero entry. Alternatively, μ_0 may be set to zero. Note that although the levels may be different, the slopes are the same, irrespective of whether they are

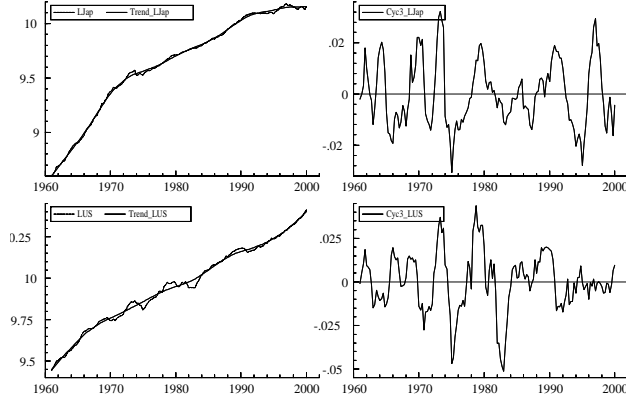


Figure 2.1: Trends and cycles from a bivariate structural time series model

fixed or stochastic. In terms of (2.7), the matrices Σ_η and Σ_ζ are of rank one, since they are $\Sigma_\eta = \sigma_\eta^2 \mathbf{ii}'$ and $\Sigma_\zeta = \sigma_\zeta^2 \mathbf{ii}'$ respectively.

A balanced growth model implies that the series have a stable relationship over time. This means that there is a full rank $(N - 1) \times N$ matrix, \mathbf{D} , with no null columns and the property that $\mathbf{D}\mathbf{i} = \mathbf{0}$, thereby rendering $\mathbf{D}\mathbf{y}_t$ jointly stationary. The rows of \mathbf{D} may be termed balanced growth co-integrating vectors. Typically each row will contain a one, a minus one and zeroes elsewhere. For example, one country may be used as a benchmark or numeraire.

If the series are stationary in first differences, balanced growth may be incorporated in a vector error correction model (VECM) by writing

$$\Delta \mathbf{y}_t = \boldsymbol{\delta} + \Gamma \mathbf{D} \mathbf{y}_{t-1} + \sum_{r=1}^p \Phi_r^* \Delta \mathbf{y}_{t-r} + \boldsymbol{\xi}_t, \quad \text{Var}(\boldsymbol{\xi}_t) = \Sigma_\xi \quad (2.11)$$

where the Φ_r^* s are $N \times N$ matrices, \mathbf{D} is as defined in the previous paragraph and the matrix Γ is $N \times (N - 1)$. The system has a single unit root, guaranteed by the fact that $\mathbf{D}\mathbf{i} = \mathbf{0}$. The constants in $\boldsymbol{\delta}$ contain information on the common slope, β , and on the differences in the levels of the series, as contained in the vector $\boldsymbol{\alpha}$. These differences might be parameterised with respect to the contrasts in $\mathbf{D}\mathbf{y}_{t-1}$. For example if $\mathbf{D}\mathbf{y}_t$ has elements $y_{it} - y_{i+1,t}$, $i = 1, \dots, N - 1$, then α_i , the i -th element of the $(N - 1) \times 1$ vector $\boldsymbol{\alpha}$, is the gap between y_i and y_{i+1} . In any case,

$\boldsymbol{\delta} = \beta(\mathbf{I} - \sum_{j=1}^p \Phi_j^*)\mathbf{i} - \Gamma\boldsymbol{\alpha}$. Estimation by OLS applied to each equation in turn is fully efficient since each equation contains the same explanatory variables.

A UC balanced growth model in which the common trend is a random walk plus drift may be approximated by (2.11); Koopman and Harvey (2001) show how to compute the coefficients. This can be useful as a baseline for forecasting and for giving initial estimates of some parameters. However, the VECM does not provide the description that can be obtained by extracting unobserved components. Furthermore, an autoregressive approximation to a UC model with a smooth trend will typically require too many lags to be viable.

2.5. Stability tests

Bivariate stability tests are testing whether two nonstationary series are evolving in such a way that their difference is stationary. Multivariate tests are testing whether all pairs have stationary differences.

The test of stability is carried out by applying the multivariate stationarity test to $\mathbf{D}\mathbf{y}_t$. The test has the rejection region

$$\eta(N-1) = \text{tr} [\mathbf{S}^{-1}\mathbf{C}] > c, \quad (2.12)$$

where

$$\mathbf{C} = T^{-2} \sum_{i=1}^T \left[\sum_{t=1}^i \mathbf{e}_t \right] \left[\sum_{t=1}^i \mathbf{e}_t \right]' \quad \text{and} \quad \mathbf{S} = T^{-1} \sum_{t=1}^T \mathbf{e}_t \mathbf{e}_t'. \quad (2.13)$$

where $\mathbf{e}_t = \mathbf{D}\mathbf{y}_t - \overline{\mathbf{D}\mathbf{y}}$. Under the null hypothesis, the limiting distribution of $\eta(N-1)$ is Cramér-von Mises with $N-1$ degrees of freedom, $CvM(N-1)$.

The test can be modified so as to include a nonparametric correction for serial correlation as in Nyblom and Harvey (2000). This is a generalisation of the univariate test of Kwiatkowski *et al* (1992). When the test statistic uses the first k sample autocovariances to compute the long-run variance it will be written $\eta(N-1; k)$. Parametric adjustments can also be made. The choice of \mathbf{D} is not crucial since pre-multiplication of $\mathbf{D}\mathbf{y}_t$ by a non-singular $(N-1) \times (N-1)$ matrix leaves the test statistic unchanged. Kuo and Mikkola (2001) argue that the multivariate η test is attractive for testing whether groups of countries exhibit purchasing power parity because it accounts for cross-sectional correlation and is invariant to the benchmark currency.

If there are no constant terms in (2.10), that is $\boldsymbol{\alpha} = \mathbf{0}$, the series contain an *identical* common trend. Stability tests can be carried out with this restriction

imposed and the distribution of the $\eta(N-1)$ statistic is $CvM_0(N-1)$; see Hobijn and Franses (2000). An LR or Wald test of the identical trends restriction, $\boldsymbol{\alpha} = \mathbf{0}$, can be readily carried out; the asymptotic distribution of the test statistic under the null hypothesis is χ^2_{N-1} . However, the test may be done without fitting a model since the variance of the mean of $\mathbf{D}\mathbf{y}_t$ - the long-run variance of \mathbf{y}_t - is $2\pi\mathbf{F}(0)$, where $\mathbf{F}(0)$ is the (multivariate) spectrum of $\mathbf{D}\mathbf{y}_t$ at frequency zero. This can be estimated nonparametrically as in the multivariate stationarity test. Such a test should logically come after the $CvM_1(N-1)$ test has indicated stability.

3. Convergence

3.1. Definitions of convergence and their implications for testing

Two countries *have converged* if the difference between them is stable. If initial conditions are unimportant, stability implies that the difference between the series, y_t , is stationary for virtually the whole period. If the mean of y_t is zero the countries are in a state of *absolute convergence*. If the mean, α , is not zero we have *conditional* or *relative convergence*; this is a possibility if we entertain the existence of increasing costs of convergence and possible barriers to absolute convergence. The limiting growth paths for the regions are then parallel, differing by α .

Although they sometimes purport to be testing whether economies are converging, most ‘convergence tests’ are actually testing stability. This is consistent with the view of Bernard and Durlauf (1996 p 170) who state that the time series approaches to convergence check for the compatibility of the difference in (log) output with an indeterministic stationary series. Furthermore Bernard and Durlauf (1996 p 171) point out ‘In time series tests, one assumes that the data are generated by economies near their limiting distributions and convergence is interpreted to mean that initial conditions have no (statistically significant) effect on the expected value of output differences.’ In most applied studies the convergence tests are unit root tests, usually (augmented) Dickey-Fuller. However, if it is stability which is being tested, stationarity tests are arguably more appropriate. It is interesting that Hobijn and Franses (2000) use stationarity tests but then add to the confusion by saying that they are testing whether the countries “*are converging*”.

If all that can be managed is stability tests, it is a pity, because this is not addressing any questions on the *process* of convergence. Data on countries often shows that they are converging, have just converged or have converged some time

ago but still have a large part of the series dependent on initial conditions. This is certainly the case with the US and Japan and the null hypothesis of stability is easily rejected even though there appears to have been convergence towards the end of the sample: $\eta(1;4) = 2.56$ and $\eta(1;8) = 1.47$, while the 5% critical value is 0.461.

Bernard and Durlauf (1996) offer a definition of ‘convergence as catching up’ over the period t to $t + \tau$. This is based on information, I_t , at time t and is

$$E(y_{t+\tau}|I_t) < y_t \quad (3.1)$$

Implementation of this definition requires a model, which can be univariate if I_t is just the series itself. However, if the model is thought of as being made up of a long-run underlying component, μ_t , and a short-run transient component, then the condition is seen to be inadequate and it may be violated because of an unusually pronounced short-run component at time t . Within such a framework a better definition of catching up over the period t to $t + \tau$ is

$$E(\mu_{t+\tau}|I_t) < E(\mu_t|I_t) \quad (3.2)$$

Catching up can be assessed after the event by defining convergence to have taken place over a particular time period if

$$E(\mu_{t+\tau}|I_T) < E(\mu_t|I_T), \quad t + \tau \leq T. \quad (3.3)$$

The next sub-section looks at how convergence in the sense of (3.3) can be assessed. The issue of whether two economies have converged absolutely can also be addressed by asking whether the estimate of $E(\mu_T|I_T)$ is significantly different from zero. The models used do not necessarily satisfy the long-run definition of convergence given in (3.4) below, though as will be seen in sub-section 3.3 there is a connection.

If a model is fitted, it will imply convergence to a fixed level, α , if

$$\lim_{\tau \rightarrow \infty} E(y_{t+\tau}|I_t) = \alpha \quad (3.4)$$

Convergence is absolute if $\alpha = 0$ as in Bernard and Durlauf (1996). If the initial condition is positive (negative), that is $\mu_0 > 0$ ($\mu_0 < 0$), we might add the requirement $\alpha < \mu_0$ ($\alpha > \mu_0$). The models set up in sub-section 3.3 are able to satisfy this condition and they become stationary for economies which have converged. Within this modelling framework it can be seen that particular forms of unit root tests are appropriate for testing hypothesis concerning the process of convergence.

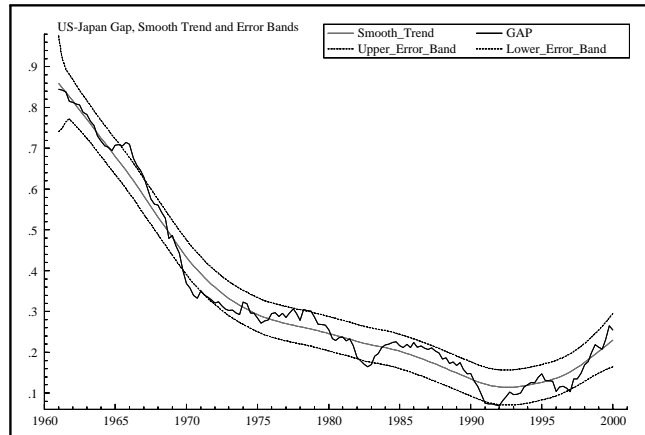


Figure 3.1: US-Japan gap modelled by a smooth trend

3.2. Direct tests based on unobserved components models

Suppose we wish to look at stylised facts without positing a particular mechanism for convergence. The difference, y_t , is assumed to be made up of a stochastic trend or level, μ_t , together with other components such as cycle and irregular as in (2.4). The smoothed estimates of the trend describe the time path reflecting the long-run difference between the two economies. Simply plotting this time path may be very informative. For example, figure 3.1 shows the difference in the trend of per capita GDP between the USA and Japan obtained by fitting a smooth trend, that is with σ_η^2 set to zero¹, plus cycle model using the STAMP package of Koopman et al (2000). We can go further and carry out tests of whether the gap between the two economies has narrowed significantly and/or whether the gap is zero, that is $\mu_T = 0$, indicating that absolute convergence has taken place. The result can be seen from the graph where a confidence interval of two RMSE's is shown. The level in the trend at the end of the sample is 0.230 with a RMSE of 0.032 giving a '*t-value*' of 7.10. Although Japan came close to catching up with the USA in the early 1990's the movement since then has been in the opposite direction.

The above test of absolute convergence was carried out simply by comparing the estimate of μ_T with its RMSE. This information is readily available from the

¹If the level variance is not set to zero, the trend reduces to a random walk plus drift and the cycle effectively disappears.

output of the Kalman filter. More generally, if there are N series we can model them all or we can model a set of $N - 1$ contrasts, $\mathbf{D}\mathbf{y}_t$, where the matrix \mathbf{D} is defined as in the sub-section on stability tests. A multivariate test of the absolute convergence hypothesis, $\mathbf{D}\boldsymbol{\mu}_T = \mathbf{0}$, can be easily constructed and treated as having an asymptotic distribution under the null which is χ_{N-1}^2 . For given values of the parameters in the model of all N series, the test will be invariant to the choice of \mathbf{D} .

Relative convergence can be assessed by testing the null of no change in the difference between two economies after a certain point. A test of this kind requires that we assess the significance of the underlying change in the series between time $t = \tau$ and time $t = T$. This requires estimating the *temporal (level) contrast*, $\mu_\tau - \mu_T$, and its root mean square error (RMSE). The RMSE is most easily obtained by setting up a fixed-point smoother to calculate the estimator of the contrast, $\tilde{\mu}_{\tau|T} - \tilde{\mu}_{T|T}$; see Harvey and Bernstein (2001). A test of significance, possibly one-sided, can be carried out by comparing the *temporal contrast test statistic*

$$\frac{\tilde{\mu}_{\tau|T} - \tilde{\mu}_{T|T}}{RMSE(\tilde{\mu}_{\tau|T} - \tilde{\mu}_{T|T})} \quad (3.5)$$

with a standard normal distribution.

Temporal contrasts can also be used to measure the extent to which there has been catching up in the sense of (3.3).

The use of non-stationary components to model convergence is apparently contradictory since once convergence has taken place the series are stationary. The next sub-section sets out stable models for μ_t instead of approximating it by a stochastic trend.

3.3. Modelling the dynamics of converging economies

An error correction mechanism (ECM) can be used to capture convergence dynamics. Time series tests of whether economies are converging can then be formulated within this framework. This reflects a view that modelling comes before testing and that testing in a vacuum is not very informative.

The simplest model is

$$y_t = \alpha + \mu_t, \quad \mu_t = \phi\mu_{t-1} + \eta_t, \quad t = 1, \dots, T, \quad (3.6)$$

with a fixed initial value, μ_0 . The crucial point is that this is not constructed as a model of a stable contrast but rather as a model of transitional dynamics in a

situation where the initial value is some way from zero. If $\phi < 1$, the gap tends to narrow over time. It makes little sense to have ϕ negative and for $0 < \phi < 1$, the model satisfies all the definitions of convergence given sub-section 3.1. Of course when the initial conditions have worked themselves out, the series becomes stationary. The equivalent error correction (EC) representation for μ_t is

$$\Delta y_t = (\phi - 1)(y_{t-1} - \alpha) + \eta_t = \delta + (\phi - 1)y_{t-1} + \eta_t, \quad t = 2, \dots, T, \quad (3.7)$$

where $\delta = \alpha(1 - \phi)$, and this can be interpreted as saying that, for data in logarithms, the expected growth rate in the current period is a negative fraction of the gap between the two economies after allowing for the permanent difference, α . For example, with $\phi = 0.98$ and a ratio of 1.65 in income per head, which corresponds to a gap in logarithms of 0.5, the difference in growth rates is 1%. Some idea of what different values of ϕ imply about the closing of the gap can be obtained by noting that the τ -step ahead forecast from an AR(1) model is ϕ^τ times the current value. Thus ϕ^τ is the fraction of the gap expected to remain after τ time periods. Table 2a shows some values of ϕ^τ .

Writing the model in EC form accords with the notion of convergence in the cross-sectional literature, as expounded by Barro and Sala-i-Martin (1992) and others, except that there the growth rate is taken to be a linear function of the initial value, giving a model which is internally inconsistent over time; see Evans and Karras (1996, p 253).

The ECM may be generalised to allow for richer dynamics. Within an autoregressive framework, (3.7) may be augmented with lagged values of differenced observations. Fitting such a model to the US-Japan series without the constant gave

$$\widehat{\Delta y}_t = -0.0086y_{t-1} + 0.127\Delta y_{t-1} + 0.083\Delta y_{t-2} + 0.136\Delta y_{t-3} + 0.128\Delta y_{t-4},$$

The equation standard error, denoted SE (equal to $\widehat{\sigma}_\eta$ here) is 0.0126 and $Q(11)$, the Box-Ljung statistic based on 11 residual autocorrelations, is 7.29². With a constant added to the right hand side

$$\widehat{\Delta y}_t = \begin{array}{cccccc} 0.0029 & -0.0156y_{t-1} & +0.118\Delta y_{t-1} & +0.076\Delta y_{t-2} & +0.133\Delta y_{t-3} & +0.127\Delta y_{t-4}, \\ (0.0019) & (0.0056) & (0.081) & (0.083) & (0.083) & (0.083) \end{array}$$

²Under the null hypothesis of correct specification, the asymptotic distribution is χ_6^2 .

with $SE = 0.0125$ and $Q(11) = 6.84$. The estimate of ϕ has fallen from 0.991 to 0.984. The ‘ t -statistic’ of the constant is 1.54 and the implied value of α is 0.187. None of the lagged differences is statistically significant at the 5% level. With no lags, the estimate of ϕ was 0.979 and the implied value of α was 0.143. However, there was evidence of residual serial correlation with $Q(11) = 25.14$.³

With a constant included, the *ADF* t -statistic is -2.77. The null hypothesis of no convergence cannot be rejected if we use the 5% (asymptotic) critical value of -2.90. This is clearly absurd if we look at figure 3.1. As Bernard and Durlauf (1996, p 172) say..‘ time series results accepting the no convergence null may be due to transitional dynamics..’. Dropping the constant, on the other hand, gives an *ADF* t -statistic of -2.57 which is significant since the (asymptotic) critical value is -1.95.

Rather than simply carrying out a test, we might wish to put a confidence interval around the rate at which convergence is taking place. The evidence in Stock (1991) indicates that such intervals could be rather wide.

3.4. Unobserved components

The UC approach is to add cycle and irregular components to the error correction mechanism. This avoids confounding the transitional dynamics of convergence with short-term steady-state dynamics. Thus

$$y_t = \alpha + \mu_t + \psi_t + \varepsilon_t, \quad \mu_t = \phi\mu_{t-1} + \eta_t, \quad t = 1, \dots, T. \quad (3.8)$$

Estimation is effected by using the state space form with a diffuse prior for μ_t (as though it were nonstationary). Although α is regarded as a fixed parameter, it can also be estimated by including it in the state vector with a diffuse prior. Care must be taken as α is not identified when ϕ is unity; one option is to carry out numerical optimisation with a transformation, such as $-\log(1-\phi)$ lying between 0 and ∞ , which keeps ϕ strictly less than one. Nevertheless, as the appendix shows, it may still be difficult to estimate α accurately. A likelihood ratio test of the null hypothesis that $\alpha = 0$ can be carried out, but in order to ensure comparability of likelihood the one for the unrestricted model must be calculated by treating α as being fixed.

³Under the null hypothesis of correct specification, the asymptotic distribution is χ_{10}^2 . The 1% significance value is 23.2

Smother transitional dynamics can be achieved by specifying μ_t in (3.8) as

$$\begin{aligned}\mu_t &= \phi\mu_{t-1} + \beta_{t-1}, \quad t = 1, \dots, T, \\ \beta_t &= \phi\beta_{t-1} + \zeta_t,\end{aligned}$$

If we write the model as what might be termed a second-order EC model,

$$\begin{aligned}\Delta\mu_t &= (\phi - 1)\mu_{t-1} + \beta_{t-1}, \quad t = 1, \dots, T, \\ \Delta\beta_t &= (\phi - 1)\beta_{t-1} + \zeta_t,\end{aligned}\tag{3.9}$$

it can be seen that there is a convergence mechanism operating on both the gap in the level and the gap in the growth rate. Alternatively this second-order ECM can be expressed as

$$\Delta\mu_t = -(1 - \phi)^2\mu_{t-1} + \phi^2\Delta\mu_{t-1} + \zeta_t$$

showing that the underlying change depends not only on the gap but also on the change in the previous time period. This means that changes take place more slowly.

The model is equivalent to an AR(2) process with both roots equal to ϕ . Obviously the condition for stationarity is $|\phi| < 1$. The variance is $[(1 + \phi^2)/(1 - \phi^2)^3]\sigma_\zeta^2$. With a value of ϕ close to one, μ_t will behave in a similar way to the smooth trend shown in figure 3.1. On the other hand, the first-order ECM behaves rather like a random walk specification and tracks the observations closely, leaving little scope for the addition of short-term non-transitional components. The ACF of the model is

$$\rho(\tau) = [1 + \{(1 - \phi^2)/(1 + \phi^2)\}\tau]\phi^\tau,$$

so the decay is slower than in an AR(1) with the same value of ϕ . For the smooth convergence model the ℓ -step ahead forecast function, standardised by dividing by the current value of the gap, is $\tilde{\mu}_{T+\ell|T} = (1 + c\ell)\phi^\ell$, $\ell = 0, 1, 2, \dots$ where c is a constant which depends on the ratio, λ , of the gap in the previous time period to the current one, that is $\lambda = \mu_{T-1}/\mu_T$. Since the one-step ahead forecast is $2\phi - \phi^2\lambda$, it follows that $c = 1 - \phi\lambda$, so

$$\tilde{\mu}_{T+\ell|T} = (1 + (1 - \phi\lambda)\ell)\phi^\ell, \quad \ell = 0, 1, 2, \dots$$

If $\lambda = 1/\phi$, the expected convergence path is the same as in the first order model. If λ is set to $2/(1 + \phi^2)$, the convergence path is the same way as the ACF. In

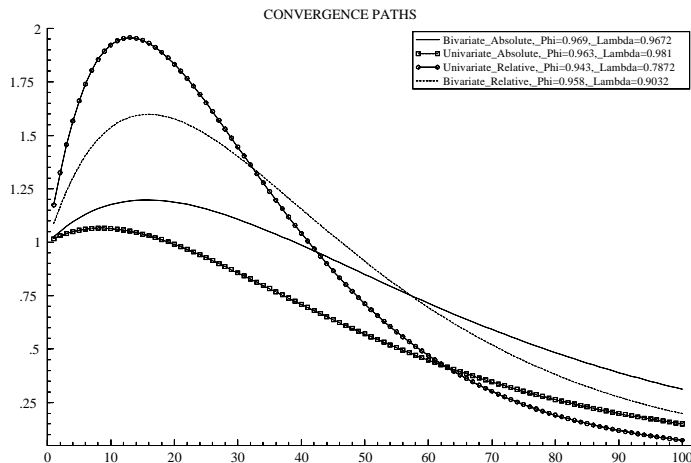


Figure 3.2: Convergence paths for models with smooth convergence dynamics

this case, the slower convergence can be illustrated by noting, for example, that with $\phi = 0.96$, 39% of the gap can be expected to remain after 50 time periods as compared with only 13% in the first-order case.

The most interesting aspect of the second-order model is that if the convergence process stalls sufficiently, the gap can be expected to widen in the short run. Figure 3.2 shows some of the paths obtained by fitting various models to data on the US and Japan; the details of the bivariate models are reported in the next section.

Estimating the first-order UC model, (3.8), resulted in relatively small values for the cycle and irregular variances. The same thing happened when a random walk trend was fitted in the preliminary model, instead of a smooth trend. The dominance of the transitional component over the cycle and irregular means that the model is not too far from a simple ECM as in (3.7). The convergence parameter, ϕ , is 0.984 for absolute convergence and 0.977 when α is estimated. The estimate of α is 0.134, but the LR statistic is 3.33 which is not significant against a χ_1^2 distribution. This is perhaps an indication of flatness of the likelihood and the associated difficulty of estimating the gap accurately.

The second-order convergence model, (3.9), fared much better insofar as it was able to separate out a cyclical component. The results are shown in table 2. The smoothed path of μ_t is very similar to that shown in figure 1. Figure 3.3 shows the predictions for the series, y_t , over a twenty year horizon. These predictions show

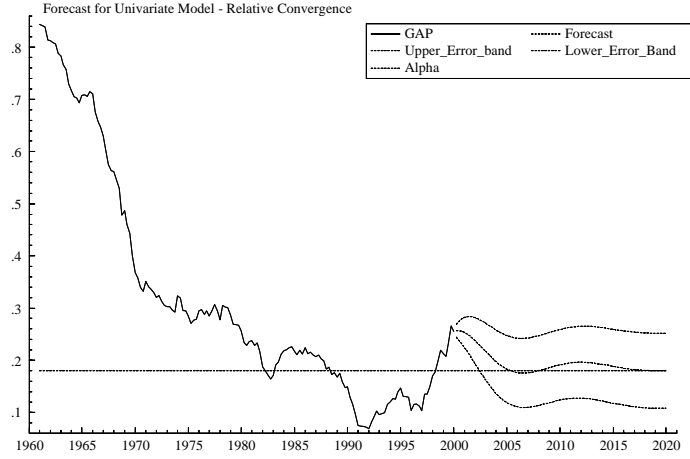


Figure 3.3: Forecasts for convergence model

some influence from the cycle, which is estimated to have a rather high period of over twelve years. The parameters obtained when the smooth trend was fitted to give figure 3.1 are shown in the last column. The estimate of α now has a statistically significant LR statistic of 6.46.

<i>TABLE 2</i>		Smooth	Convergence	Model
<i>Hyperparameters</i>		Absolute	Relative	Trend
<i>Convergence</i>	$\sigma_{\zeta}(\times 10^{-3})$	1.933	1.286	1.244
	ϕ	0.963	0.943	1 (fixed)
<i>Cycle</i>	$\sigma_{\kappa}(\times 10^{-3})$	11.33	11.51	11.25
	ρ_c	0.94	0.96	0.95
	<i>Period</i> ($2\pi/\lambda_c$)	50.51	50.14	50.91
<i>Irregular</i>	$\sigma_{\varepsilon}(\times 10^{-3})$	0.014	0.071	1.54
<i>Gap</i>	α	0 (fixed)	0.180	—
<i>Fit</i>	$\log L$	454.679	457.909	451.94
	$SE(\times 10^{-3})$	12.7	12.4	12.8
<i>Diagnostics</i>	$Q(11)$	11.54	10.85	9.37

4. Bivariate models for the levels of converging economies

The previous section devised a mechanism for capturing convergence between two economies. This section explores how this mechanism can be incorporated within a bivariate model for the levels of converging economies. The aim is to be able to extract trend and convergence components and to formulate forecasts which take convergence to a common trend into account.

4.1. Bivariate error correction mechanism

A bivariate model for two converging economies can be set up as

$$\begin{aligned}\Delta y_{1t} &= \phi_1(y_{2,t-1} - y_{1,t-1}) + \eta_{1t} \\ \Delta y_{2t} &= \phi_2(y_{1,t-1} - y_{2,t-1}) + \eta_{2t}\end{aligned}\tag{4.1}$$

where $y_{i,t}$ denotes, for example, per capita output for economy i at time t . Absolute convergence and no growth is initially assumed for simplicity. Thus the growth rate of the first economy depends on the gap between its level and that of the second economy and vice versa.

The model corresponds to the first-order vector autoregression

$$\begin{aligned}y_{1t} &= (1 - \phi_1)y_{1,t-1} + \phi_1 y_{2,t-1} + \eta_{1t} \\ y_{2t} &= \phi_2 y_{1,t-1} + (1 - \phi_2)y_{2,t-1} + \eta_{2t}\end{aligned}\tag{4.2}$$

The roots of the transition matrix

$$\Phi = \begin{bmatrix} 1 - \phi_1 & \phi_1 \\ \phi_2 & 1 - \phi_2 \end{bmatrix}$$

are one and $1 - \phi_1 - \phi_2$. The condition for the second root to lie inside the unit circle is $0 < \phi_1 + \phi_2 < 2$. This being the case, the long-run forecasts converge to the same value, that is $\bar{\phi}y_{1T} + (1 - \bar{\phi})y_{2T}$, since

$$\lim_{k \rightarrow \infty} \Phi^k = \begin{bmatrix} \bar{\phi} & 1 - \bar{\phi} \\ \bar{\phi} & 1 - \bar{\phi} \end{bmatrix}\tag{4.3}$$

where $\bar{\phi} = \phi_2/(\phi_1 + \phi_2)$. This is a standard result in the theory of Markov chains; see next section.

The model (4.2) can be premultiplied by a matrix with unit Jacobian thereby transforming it to

$$y_{1t} - y_{2t} = \phi(y_{1,t-1} - y_{2,t-1}) + \eta_{1t} - \eta_{2t}$$

$$\bar{y}_{\phi t} = \bar{y}_{\phi,t-1} + \bar{\eta}_{\phi t}$$

where $\phi = 1 - (\phi_1 + \phi_2)$ and

$$\bar{y}_{\phi t} = \bar{\phi}y_{1t} + (1 - \bar{\phi})y_{2t} \quad (4.4)$$

with the disturbance $\bar{\eta}_{\phi t}$ defined similarly. The first equation is a convergence mechanism for the difference $y_{1t} - y_{2t}$. In the second equation the weighted sum follows a random walk and, as is clear from (4.3), this is the growth path to which the two economies are converging.

Parameterising the model in terms of $\bar{\phi}$ and ϕ has some attractions. The stability condition is $|\phi| < 1$, though it makes little sense to have ϕ negative. It seems desirable (though not essential for stability) to have $0 \leq \bar{\phi} \leq 1$. This condition implies that ϕ_1 and ϕ_2 are both greater than or equal to zero. Note that if $\phi = 1$, then $\bar{\phi}$ is not identified. At first sight there may seem to be a contradiction in the two parameterisations in that there are two rates of convergence in (4.2). However, these rates refer to the change in each economy rather than to the change in the gap.

Benchmark model Setting $\phi_2 = 0$ (or $\phi_1 = 0$) implies that country one (two) converges to country two (one), the benchmark country. Provided ϕ_1 is positive, $\phi_2 = 0$ does not imply a second unit root and so a test of this hypothesis can be based on standard distribution theory. Note that $y_{1,t-1} - y_{2,t-1}$ is stationary (the variables are co-integrated) in (4.1).

Trend and constant The model may be extended to include a common deterministic trend and a constant, α , to allow for relative convergence. Thus

$$\begin{aligned} y_{1t} &= \alpha + \beta t + \mu_{1t} \\ y_{2t} &= \beta t + \mu_{2t} \end{aligned}$$

where

$$\begin{aligned} \Delta\mu_{1t} &= \phi_1(\mu_{2,t-1} - \mu_{1,t-1}) + \eta_{1t} \\ \Delta\mu_{2t} &= \phi_2(\mu_{1,t-1} - \mu_{2,t-1}) + \eta_{2t} \end{aligned} \quad (4.5)$$

The gap, $y_{2t} - y_{1t}$, is as in (3.6), except that the sign of α is different (this is more convenient for what follows). Substituting for μ_{1t} and μ_{2t} gives

$$\begin{aligned}\Delta y_{1t} &= \beta - \phi_1 \alpha + \phi_1 (y_{2,t-1} - y_{1,t-1}) + \eta_{1t} \\ \Delta y_{2t} &= \beta + \phi_2 \alpha + \phi_2 (y_{1,t-1} - y_{2,t-1}) + \eta_{2t}\end{aligned}\tag{4.6}$$

Note that the weighted average, (4.4), is a random walk with a drift of β and that the gap, $y_{2t} - y_{1t}$, satisfies the ECM of (3.7).

4.2. Autoregressive model

The dynamics in (4.6) may be extended by adding lagged differences to the right hand side of the equations and re-arranging to give

$$\begin{aligned}\Delta y_{1t} &= \delta_1 - \phi_1 (y_{1,t-1} - y_{2,t-1}) + \sum_{r=1}^p \phi_{11r}^\dagger \Delta y_{1,t-r} + \sum_{r=1}^p \phi_{12r}^\dagger \Delta y_{2,t-r} + \eta_{1t} \\ \Delta y_{2t} &= \delta_2 + \phi_2 (y_{1,t-1} - y_{2,t-1}) + \sum_{r=1}^p \phi_{21r}^\dagger \Delta y_{1,t-r} + \sum_{r=1}^p \phi_{22r}^\dagger \Delta y_{2,t-r} + \eta_{2t}\end{aligned}\tag{4.7}$$

where $\delta_i = \beta(1 - \sum_{j=1}^p (\phi_{i1j}^\dagger + \phi_{i2j}^\dagger)) + (-1)^i \phi_i \alpha$, $i = 1, 2$. The parameters α and β can be identified from the estimated constants once estimates of ϕ_1 and ϕ_2 have been obtained. The model belongs to the VECM class of (2.11). The co-integrating vector is known and ML estimation can be carried out by OLS since the regressors are the same in each equation. If we were to set α to zero then the restriction that the slopes are the same would need to be enforced.

In the benchmark model, ϕ_i is set to zero in one equation and so β is identified from that equation. Using the estimate of β , an estimate of α can be extracted from the estimated constant in the other equation. There should, in theory, be gains from SURE estimation, although in practice it seems to make little difference here.

A bivariate model was estimated for the US and Japan with $p = 4$. For Japan we find $\tilde{\phi}_1 = 0.0184$ while for the US, $\tilde{\phi}_2 = -0.0046$. The model is stable, but the negative sign for $\tilde{\phi}_2$ suggests that it should be set to zero, as in a benchmark model. Indeed the ‘ t -statistic’ is only 0.937; recall that this is asymptotically standard normal provided ϕ_1 is positive. The benchmark model gave an estimate of ϕ_1 equal to 0.0176, corresponding to $\phi = 0.9824$. From the estimates of the constants, α and β are estimated as 0.140 and 0.0062 respectively. The estimate of β corresponds to an annual growth rate of 2.6%. Recall that the univariate estimate of α from modelling the difference as an autoregression was 0.143.

4.3. Unobserved components

Embedding the ECM within a UC model by adding a cycle and an irregular to (4.5) gives

$$\begin{aligned} y_{1t} &= \alpha + \beta t + \mu_{1t} + \psi_{1t} + \varepsilon_{1t} \\ y_{2t} &= \beta t + \mu_{2t} + \psi_{2t} + \varepsilon_{2t} \end{aligned} \quad (4.8)$$

If ψ_{1t} and ψ_{2t} are modelled as similar cycles, as in (2.8), subtracting y_{1t} from y_{2t} in (4.8) gives a univariate model of the form (3.8).

The vector $(\mu_{1t}, \mu_{2t})'$ may be initialised with a diffuse prior in the SSF. The parameters α and β may also be included in the state and initialised with a diffuse prior, though in order to compare likelihoods they should be treated as fixed. Note that if $\phi_1 = \phi_2 = 0$, then there is no convergence. The pure trend model of section 2 is obtained provided $\alpha = 0$. However, the balanced growth model of (2.10) is obtained if $\eta_{1,t}$ and $\eta_{2,t}$ are perfectly correlated with the same variance.

The deterministic trend could be generalised so as to become a smooth stochastic trend, that is βt is replaced by μ_t^β where $\Delta^2 \mu_t^\beta = \zeta_t$. There is no confounding with the common level of μ_{1t} and μ_{2t} provided that the initial value of the common trend level, μ_0^c , is set to zero or the initial value of either μ_{1t} or μ_{2t} is set to zero. However, the introduction of the smooth trend is more straightforward conceptually if a second-order model for the convergence dynamics, generalising (3.9), is set up. We then have

$$\begin{aligned} y_{1t} &= \alpha + \mu_{1t} + \psi_{1t} + \varepsilon_{1t}, \\ y_{2t} &= \mu_{2t} + \psi_{2t} + \varepsilon_{2t}, \end{aligned} \quad (4.9)$$

$$\begin{aligned} \mu_{1t} &= (1 - \phi_1)\mu_{1,t-1} + \phi_1\mu_{2,t-1} + \beta_{1,t-1}, \\ \beta_{1t} &= (1 - \phi_1)\beta_{1,t-1} + \phi_1\beta_{2,t-1} + \zeta_{1,t}, \\ \mu_{2t} &= (1 - \phi_2)\mu_{2,t-1} + \phi_2\mu_{1,t-1} + \beta_{2,t-1}, \\ \beta_{2t} &= (1 - \phi_2)\beta_{2,t-1} + \phi_2\beta_{1,t-1} + \zeta_{2,t}. \end{aligned} \quad (4.10)$$

Again, if $\phi_1 = \phi_2 = 0$, then there is no convergence but a balanced growth model is obtained if $\zeta_{1,t}$ and $\zeta_{2,t}$ are perfectly correlated with the same variance.

The model can be re-arranged so as to have two convergence components defined in terms of deviations from the common trend, that is $\mu_{it}^\dagger = \mu_{it} - \bar{\mu}_{\phi t}$ and $\beta_{i,t}^\dagger = \beta_{it} - \bar{\beta}_{\phi t}$, $i = 1, 2$, where $\bar{\mu}_{\phi t}$ is

$$\bar{\mu}_{\phi t} = \bar{\phi}\mu_{1t} + (1 - \bar{\phi})\mu_{2t}, \quad (4.11)$$

and similarly for $\bar{\beta}_{\phi t}$. Then

$$\begin{aligned} y_{1t} &= \alpha(1 - \bar{\phi}) + \mu_{1t}^\dagger + \mu_{\phi t} + \psi_{1t} + \varepsilon_{1t}, \\ y_{2t} &= -\bar{\phi}\alpha + [-\bar{\phi}/(1 - \bar{\phi})]\mu_{1t}^\dagger + \mu_{\phi t} + \psi_{2t} + \varepsilon_{2t}, \end{aligned}$$

with

$$\begin{aligned} \mu_{1t}^\dagger &= \phi\mu_{1,t-1}^\dagger + \beta_{1,t-1}^\dagger, \\ \beta_{1t}^\dagger &= \phi\beta_{1,t-1}^\dagger + \zeta_{1,t}^\dagger, \\ \mu_{\phi t} &= \mu_{\phi,t-1} + \beta_{\phi,t-1}, \\ \beta_{\phi t} &= \beta_{\phi,t-1} + \zeta_{\phi,t}, \end{aligned}$$

The second convergence component can be obtained from the first since $\bar{\phi}\mu_{1t}^\dagger + (1 - \bar{\phi})\mu_{2t}^\dagger = 0$. Both economies converge to the growth path of the common trend, except insofar as the first economy is at a constant level, α , above (or below) the second one. Convergence is at the same rate, ϕ .

If the second economy is taken to be a benchmark then $\phi_2 = 0$ in the last two equations of (4.10). In this case μ_{2t} is a smooth trend. Setting up the model as in (4.12) with $\bar{\phi} = 0$, focusses attention on the transitional gap between the two economies as $\mu_{1t}^\dagger = \mu_{1t} - \mu_{2t}$ and $\beta_{1,t}^\dagger = \beta_{1,t} - \beta_{2,t}$. The implied model for $y_{1t} - y_{2t}$ is as in (3.8) with μ_t replaced by μ_{1t}^\dagger .

4.4. UC Model for Japan and US

The smooth stochastic trends fitted in section 2 give an indication of the kind of results which might be expected from a convergence model and they can provide starting values for some of the parameters. As already noted, it is a limiting case which results when $\phi_1 = \phi_2 = 0$ and $\alpha = 0$.

Table 4.1

	Hyperparameters	Absolute		Relative	
		Japan	US	Japan	US
Convergence	$\sigma_\zeta(\times 10^{-3})$	1.466	0.989	1.399	1.007
	ϕ	0.969		0.958	
Cycle	$\sigma_\kappa(\times 10^{-3})$	6.932	7.464	9.413	7.681
	ρ_c	0.903		0.892	
	$Period(2\pi/\lambda_c)$	24.77		28.67	
Irregular	$\sigma_\varepsilon(\times 10^{-3})$	4.482	0.521	0	0
Gap	α	0(<i>fixed</i>)		-0.174	
Fit	$\log L$	988.050		988.766	
Diagnostics	$SE(\times 10^{-3})$	8.9	10.6	8.9	10.6
	$Q(11)$	12.37	14.77	12.37	14.77

The results of fitting the bivariate convergence model are shown in table 4.1. The model was estimated with the US taken as the benchmark, with α set to zero and α unrestricted. When the more general model with no restrictions on ϕ_1 and ϕ_2 was estimated it collapsed to the benchmark model. This is consistent with what was found when the autoregressive model, (4.7), was fitted.

The main features are:

i) The cycle parameters are similar to those obtained with the bivariate pure trend model reported in table 2.1 and the fitted cycles seem to provide a more satisfactory decomposition than was obtained for the univariate model for the difference.

ii) The estimate of α is of a similar order of magnitude to the one obtained in the univariate gap model, but it has a smaller estimated standard error, 0.047. Again there is clear evidence of relative convergence, though the LR statistic is only 1.432.

iii) The estimated convergence component, μ_{1t}^\dagger , assigned to Japan, is very similar to the smoothed gap shown in Figure 3.

iv) Figure 4 shows the forecasts for the two countries. It can be seen that they converge to the same growth path, μ_t , but at a constant distance, α , apart. A value of $\alpha = -0.174$ implies that the Japanese level is about 16% below that of the US. The forecasts from the bivariate ECM will also converge to parallel paths a similar distance apart.

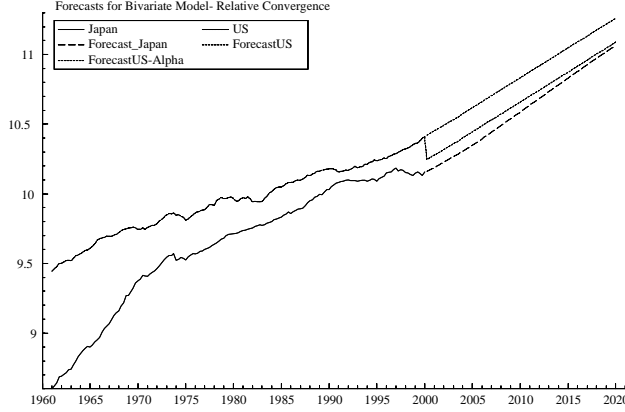


Figure 4.1:

5. Multivariate convergence models

The basic multivariate convergence model, allowing for relative convergence and a common time trend, is

$$y_{it} = \alpha_i + \beta t + \mu_{it}, \quad i = 1, \dots, N$$

with

$$\mu_{it} = \sum_{j=1}^N \phi_{ij} \mu_{j,t-1} + \eta_{it}, \quad i = 1, \dots, N \quad (5.1)$$

with $\sum_{j=1}^N \phi_{ij} = 1$ for $i = 1, \dots, N$. This restriction can be conveniently imposed by setting $\phi_{ii} = 1 - \sum_{j \neq i} \phi_{ij}$. As will be shown shortly, it ensures that the system contains a unit root. As the model stands there are $N(N-1)$ parameters governing convergence. Re-formulating it as

$$\Delta \mu_{it} = \sum_{j \neq i} \phi_{ij} (\mu_{j,t-1} - \mu_{i,t-1}) + \eta_{it}, \quad i = 1, \dots, N, \quad (5.2)$$

shows how the convergence of the i -th economy depends on the gap between it and each of the other $N-1$ economies. However, the conditions needed for stability, that is for convergence to take place, are not obvious.

In matrix form

$$\begin{aligned}\mathbf{y}_t &= \boldsymbol{\alpha} + \beta \mathbf{i}t + \boldsymbol{\mu}_t, \\ \boldsymbol{\mu}_t &= \Phi \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t\end{aligned}$$

with

$$\boldsymbol{\mu}_t = \Phi \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \quad \text{Var}(\boldsymbol{\eta}_t) = \Sigma_\eta \quad (5.3)$$

or

$$\Delta \boldsymbol{\mu}_t = (\Phi - \mathbf{I}) \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \quad \text{Var}(\boldsymbol{\eta}_t) = \Sigma_\eta$$

Since each row of Φ sums to unity, $\Phi \mathbf{i} = \mathbf{i}$. Thus setting λ to one in $(\Phi - \lambda \mathbf{I}) \mathbf{i} = \mathbf{0}$, shows that Φ has an eigenvalue of one with a corresponding eigenvector consisting of ones. The other roots of Φ are obtained by solving $|\Phi - \lambda \mathbf{I}| = 0$; they should have modulus less than one for convergence.

If we write

$$\bar{\boldsymbol{\phi}}' \Delta \boldsymbol{\mu}_t = \bar{\boldsymbol{\phi}}' (\Phi - \mathbf{I}) \boldsymbol{\mu}_{t-1} + \bar{\boldsymbol{\phi}}' \boldsymbol{\eta}_t$$

it is clear that the $N \times 1$ vector of weights, $\bar{\boldsymbol{\phi}}$, which gives a random walk must be such that $\bar{\boldsymbol{\phi}}' (\Phi - \mathbf{I}) = \mathbf{0}'$. Since the roots of Φ' are the same as those of Φ , it follows from writing $(\Phi' - \mathbf{I}) \bar{\boldsymbol{\phi}} = \mathbf{0}$ that $\bar{\boldsymbol{\phi}}$ is the eigenvector of Φ' corresponding to its unit root. This random walk, $\bar{\mu}_{\phi t} = \bar{\boldsymbol{\phi}}' \boldsymbol{\mu}_t$, is a common trend in the sense that it yields the common growth path to which all the economies converge. This is because $\lim_{j \rightarrow \infty} \Phi^j = \mathbf{i} \bar{\boldsymbol{\phi}}'$; the proof follows along the same lines as that for a well-known result on ergodic Markov chains as given, for example, in Hamilton (1994, p681).

The common trend for the observations is a random walk with drift, β , and with $\boldsymbol{\alpha}' \bar{\boldsymbol{\phi}} = 0$ each element of $\boldsymbol{\alpha}$ is a deviation from the common trend.

UC models The general formulation of the first-order UC model has

$$y_{it} = \alpha_i + \beta t + \mu_{it} + \psi_{it} + \varepsilon_{it}, \quad i = 1, \dots, N \quad (5.4)$$

together with (5.1). In matrix terms, using the notation of (2.7),

$$\mathbf{y}_t = \boldsymbol{\alpha} + \beta \mathbf{i}t + \boldsymbol{\mu}_t + \boldsymbol{\psi}_t + \boldsymbol{\varepsilon}_t,$$

with $\boldsymbol{\mu}_t$ as in (5.3).

The smooth convergence model is

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\alpha} + \boldsymbol{\mu}_t + \boldsymbol{\psi}_t + \boldsymbol{\varepsilon}_t, \\ \boldsymbol{\mu}_t &= \Phi \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}_{t-1} \\ \boldsymbol{\beta}_t &= \Phi \boldsymbol{\beta}_{t-1} + \boldsymbol{\zeta}_t \end{aligned}$$

The forecasts converge to those of the smooth common trend, $\bar{\mu}_{\phi t} = \bar{\boldsymbol{\phi}}' \boldsymbol{\mu}_t$

For both types of UC model, an LR test of the absolute convergence hypothesis, that is $\boldsymbol{\alpha} = 0$, is straightforward with the asymptotic distribution under the null being χ_{N-1}^2 .

VECM Convergence may be captured by the common trend VECM of (2.11). The matrix Γ contains $N(N-1)$ free parameters and these may be estimated by OLS applied to each equation in turn. The Φ matrix of (5.1) may then be estimated⁴ as it is given by $\Gamma \mathbf{D} + \mathbf{I}$. However, there is no guarantee that the estimate of Γ will be such $N-1$ of the roots of Φ have modulus less than one. If the vector of gaps, $\boldsymbol{\alpha}$, is parameterised with respect to the contrasts in \mathbf{Dy}_{t-1} , then $\boldsymbol{\delta} = \beta(\mathbf{I} - \sum_{j=1}^p \Phi_j^*) \mathbf{i} - \Gamma \boldsymbol{\alpha}$.

5.1. Deviation and benchmark restrictions

If there is symmetry so that $\phi_{ij} = \phi_{ji}$, the number of parameters is reduced to $N(N-1)/2$ and the mean, $\bar{\mu}_t$, is a random walk. This follows immediately from the matrix formulation as Φ is symmetric - hence $\bar{\boldsymbol{\phi}} = \mathbf{i}$. However, the restriction is not particularly appealing and there is still a large number of parameters.

A better way in which to impose restrictions is to specify a model in terms of deviations from a weighted average. This happens naturally if deviations (of per capita income) in regions from a national average are to be considered. Typically some regions will be bigger than others and so will receive more weight in constructing the average. However, in a more general situation we might consider the weights as giving some indication of influence. Let

$$\bar{\mu}_{wt} = \sum_{i=1}^N w_i \mu_{it}, \quad \sum_{i=1}^N w_i = 1$$

⁴We can also directly adopt the parameterisation implicit in the Φ matrix. Although this implies a different set of explanatory variables in each equation, all satisfy the co-integrating constraints and so OLS is efficient for each equation in turn.

and set

$$\phi_{ii} = \pi_i + 1 - \pi_i w_i = \pi_i(1 - w_i) + 1 \quad \text{and} \quad \phi_{ij} = -\pi_i w_j, \quad i \neq j. \quad (5.5)$$

Substituting in (5.2) yields

$$\begin{aligned} \Delta\mu_{it} &= \pi_i \sum_{j \neq i} w_j (\mu_{i,t-1} - \mu_{j,t-1}) + \eta_{it} \\ &= \pi_i (\mu_{i,t-1} - \bar{\mu}_{w,t-1}) + \eta_{it}, \quad i = 1, \dots, N. \end{aligned} \quad (5.6)$$

Thus $\Delta\mu_{it}$ depends on the gap between its own level and that of the weighted average. If $\pi_i = 0$, then μ_{it} is a random walk. This might be taken to suggest that having all the π_i 's negative is necessary for convergence, but as can be seen from the bivariate model this is not the case.

If $\bar{\mu}_{w,t}$ is to be a random walk, then the weights must be such that $\mathbf{w}'\Phi = \mathbf{w}'$, that is $\sum_i \phi_{ij} w_i = w_j, j = 1, \dots, N$. The weights will only satisfy this condition if, for $i = 1, \dots, N$, $\pi_i = \pi$ for $w_i \neq 0$. This being the case we have $\mathbf{w} = \bar{\phi}$.

In the homogeneous model, when all weights are non-zero and $\pi_i = \pi$ for $i = 1, \dots, N$, we are able to express the model in deviation form,

$$\Delta(\mu_{it} - \bar{\mu}_{\phi,t}) = \pi(\mu_{i,t-1} - \bar{\mu}_{\phi,t-1}) + \eta_{it} - \bar{\eta}_{\phi,t}, \quad i = 1, \dots, N, \quad (5.7)$$

and any $N - 1$ of these equations may be combined with the equation for $\bar{\mu}_{\phi,t}$ to give a complete system. The stability condition⁵ is $-2 < \pi < 0$.

There are a number of ways to proceed. If both π_i 's and w_j 's are treated as parameters, the model has $2N - 1$ parameters for $N > 2$. For moderate size N this parameterisation is relatively parsimonious. However, it is more appealing to focus attention on either the π_i 's or the w_j 's. If we let the π_i 's be the same, we can estimate the w_j 's as $\bar{\phi}_j$'s. Including π , this makes N free parameters in all. The convergence process is therefore parameterised as $\phi_{ij} = -\pi\bar{\phi}_j, i \neq j$, and $\phi_{ii} = \pi + 1 - \pi\bar{\phi}_i$. Alternatively, we may decide to pre-assign values to the w_j 's and estimate the N π_i 's. For the case $N = 2$, these two options are equivalent.

⁵The matrix \mathbf{iw}' is idempotent (though not symmetric) as its rows are identical and sum to one. Since its trace is one, it has one root of unity, while the rest are zero. The matrix $\Phi = (\mathbf{1} + \pi)\mathbf{I} - \pi\mathbf{iw}'$ also has a single unit root while the rest are $1 + \pi$.

When the π'_i 's are different, we can always calculate the implied weights, $\bar{\phi}_i$, for the common trend.

The two approaches are mixed if we set $w_i = 0$ for some i 's, and let the corresponding π'_i 's be free. If n are set to zero, we then, for $n < N - 1$, have $N - n - 1$ w'_i 's to estimate, together with n π'_i 's and one π . When $n = N - 1$, the benchmark model is obtained. In these cases, $w_j = \bar{\phi}_j$ and the μ'_{it} 's may be put in deviation form.

Deviations from the mean If we set $w_i = 1/N$, then $\bar{y}_{w,t}$ is the simple mean. The implied weights may be found from the π'_i 's since, provided $\pi_j < 0$ for all j , $\bar{\phi}_i = (1/\pi_i) / \sum_j (1/\pi_j)$ as is easily seen⁶ from (5.6). If we regard it reasonable to have $0 \leq \bar{\phi}_i \leq 1$, then the π'_i 's must be less than or equal to zero. If a $\pi_i = 0$, then we get a benchmark model, as $\bar{\phi}_{ii} \rightarrow 1$ as $\pi_i \rightarrow 0$. Within the context of (5.6) a test of $\pi_i = 0$ can be based on standard distribution theory as $\pi_i = 0$ does not, in itself, imply a unit root.

Benchmark model Take (without loss of generality) the N th country as the benchmark to which all the countries converge. Then

$$\begin{aligned}\Delta\mu_{it} &= \pi_i(\mu_{i,t-1} - \mu_{N,t-1}) + \eta_{it}, \quad i = 1, \dots, N - 1, \\ \Delta\mu_{Nt} &= \eta_{Nt},\end{aligned}$$

where the roots of the transition matrix are one and $\pi_i + 1, i = 1, \dots, N - 1$ so $-2 < \pi_i < 0, i = 1, \dots, N - 1$ for convergence. Note that this model is a special case of (5.6) obtained by setting all the weights apart from w_N equal to zero. Since μ_{Nt} is a random walk, we have

$$\Delta(\mu_{it} - \mu_{N,t}) = \pi_i(\mu_{i,t-1} - \mu_{N,t-1}) + \eta_{it} - \eta_{Nt}, \quad i = 1, \dots, N - 1. \quad (5.8)$$

A further complication with the deviation model is that if logarithms have been taken to get the y'_{it} 's, then $\bar{y}_{w,t}$ will not be the same as the the logarithm of the weighted sum of the original observations. Working in logarithms has no implications for the benchmark model.

⁶In matrix terms, $\Phi = \mathbf{I} + \Pi_D - \boldsymbol{\pi}\mathbf{i}'$, where Π_D is a diagonal matrix with the elements of $\boldsymbol{\pi}$ on its diagonal. We want to find $\bar{\boldsymbol{\phi}}$ such that $\Phi'\bar{\boldsymbol{\phi}} = (\mathbf{I} + \Pi_D - \mathbf{i}\boldsymbol{\pi}')\bar{\boldsymbol{\phi}} = \bar{\boldsymbol{\phi}}$. This can be done by making the i -th element of $\bar{\boldsymbol{\phi}}$ proportional to the inverse of the i -th element of $\boldsymbol{\pi}$. We need to standardise so that the elements sum to one.

5.2. Autoregressive models

The deviation and benchmark constraints on (5.3) can be incorporated into an autoregressive model because $y_{i,t-1} - \bar{y}_{w,t-1}$, $i = 1, \dots, N$ are all co-integrating vectors. If one is dropped it can be reconstructed as a linear combination of the others. Thus the Γ and \mathbf{D} matrices in (2.11) can be formed with suitable constraints. However, it is more convenient to set up the model as

$$\Delta y_{it} = \delta_i + \pi_i(y_{i,t-1} - \bar{y}_{w,t-1}) + \sum_{j=1}^N \sum_{r=1}^p \phi_{ijr}^* \Delta y_{j,t-r} + \eta_{it}, \quad i = 1, \dots, N, \quad (5.9)$$

The parameters may be efficiently⁷ estimated by SURE, although little is likely to be lost from simply doing OLS and this may be preferable if N is large. From the estimates of the δ'_i 's we can solve to get β and a set of α'_i 's for relative convergence since

$$\delta_i = \beta \left[1 - \sum_{j=1}^N \sum_{r=1}^p \phi_{ij}^* \right] - \pi_i \alpha_i, \quad i = 1, \dots, N. \quad (5.10)$$

In a benchmark model with $\pi_N = 0$, there are $N - 1$ gaps represented by α_i , $i = 1, \dots, N - 1$. More generally if we want them to be in terms of deviations from the level of the common trend, they must satisfy $\sum_i \bar{\phi}_i \alpha_i = 0$. Recall that with a simple mean, $\bar{\phi}_i$ is proportional to $1/\pi_i$, if all π_i 's are negative, so that the equation $\sum_i \alpha_i/\pi_i = 0$ can be added to those in (5.10).

As N becomes large, the above AR model runs into difficulties because of the potentially large number of ϕ_{ijr}^* parameters, $N^2 p$ in all.

5.3. Unobserved components model

In the autoregressive framework, the natural way to proceed when the restrictions in (5.5) are imposed is to estimate π'_i 's for a given set of pre-assigned w'_i 's. An unobserved components formulation, however, requires nonlinear optimisation with respect to the elements of Φ . Since it is unclear what constraints should be imposed on the π'_i 's, it is relatively more attractive to assume that $\pi_i = \pi$ and to estimate the $\bar{\phi}_i$ parameters, constraining them to lie between zero and one and to sum to one. The model can be transformed so as to consist of convergence

⁷If the general model, (2.11), can be estimated, a LR test of the constraints implied by (5.9) can be carried out.

processes, $\mu_{it}^\dagger, i = 1, \dots, N$, which are deviations from the common trend, $\bar{\mu}_{\phi,t}$. Then

$$y_{it} = \alpha_i + \bar{\mu}_{\phi,t} + \mu_{it}^\dagger + \psi_{it} + \varepsilon_{it}, \quad i = 1, \dots, N, \quad (5.11)$$

where $\mu_{it}^\dagger = \mu_{it} - \bar{\mu}_{\phi,t}$, with

$$\mu_{it}^\dagger = \phi \mu_{i,t-1}^\dagger + \eta_{it}^\dagger, \quad i = 1, \dots, N-1, \quad |\phi| < 1 \quad (5.12)$$

where $\phi = 1 + \pi$, the initial conditions are $\mu_{i0}^\dagger, i = 1, \dots, N-1$, are fixed, and

$$\bar{\mu}_{\phi t} = \bar{\mu}_{\phi t-1} + \beta + \bar{\eta}_{\phi t}.$$

The N -th economy has been arbitrarily omitted from the convergence equations in (5.12). It is constructed from the others as

$$\mu_{Nt}^\dagger = -\bar{\phi}_N^{-1} \sum_{i=1}^{N-1} \bar{\phi}_i \mu_{it}^\dagger \quad \text{with} \quad \bar{\phi}_N = 1 - \sum_{i=1}^{N-1} \bar{\phi}_i.$$

Similarly $\alpha_N = -\bar{\phi}_N^{-1} \sum_{i=1}^{N-1} \bar{\phi}_i \alpha_i$. These expressions can be substituted directly into the measurement equation for y_{Nt} in (5.11) above. The extension to smooth convergence processes is immediately apparent from (4.12).

In the benchmark model, $\bar{\mu}_{\phi t}$ is replaced by μ_{Nt} and we can have different π'_i s in (5.12).

6. Conclusions

The paper has set out models for converging economies and explored their implications. The main value of these models lies in the way in which they can present stylised facts and make coherent forecasts which allow for the way in which convergence is taking place.

A second-order error correction mechanism allows unobserved components models to give an informative decomposition into trend, cycle and convergence components. Univariate and multivariate models can be constructed in this way, and the issue of absolute and relative convergence may be investigated. Fitting these models to Japanese and US GDP per capita indicates that, by combining the information in the two series, the bivariate model can give a better decomposition. The more general multivariate models offer the opportunity for further analysis in this direction.

ACKNOWLEDGEMENTS

Earlier versions of this paper were presented at seminars at Nuffield College, Oxford and Erasmus University, Rotterdam and at conferences in Brasilia and Santiago, Chile. We are grateful to our colleagues at Cambridge for many helpful discussions and to David Hendry for his comments. The work was supported by the ESRC as part of a project on Dynamic Factor Analysis, grant number L138 25 1008.

A. Appendix: Estimation of a relative convergence model

For model (3.6), the log-likelihood is, excluding constants,

$$\log L = -\frac{T}{2} \log \sigma^2 - \frac{1}{2} \sum_{t=1}^T \{y_t - \phi y_{t-1} - \alpha(1 - \phi)\}^2$$

where y_0 , like α , is treated as a fixed parameter. Then

$$\tilde{y}_0 = \tilde{\alpha} + (y_1 - \tilde{\alpha}) / \tilde{\phi}$$

$$\tilde{\phi} = \frac{\sum_{t=2}^T (y_{t-1} - \tilde{\alpha})(y_t - \tilde{\alpha})}{\sum_{t=2}^T (y_{t-1} - \tilde{\alpha})^2}$$

and

$$\tilde{\alpha} = \frac{T}{T-1} \bar{y} + \frac{\tilde{\phi} y_T - y_1}{(T-1)(1 - \tilde{\phi})}$$

where $\bar{y} = \sum_{t=1}^T y_t$. If $\tilde{\phi}$ is close to one the second term on the right will be large if the first and last observations are very different.

Estimating the parameters by a regression based on (3.7) gives similar, but not identical, results and again the estimator of α is very sensitive to the estimator of ϕ .

REFERENCES

- Barro, R.J. and Sala-i-Martin (1992). Convergence. *Journal of Political Economy*, **100**, 223-51.
- Bernard, A.B. and S. Durlauf. (1996). Interpreting tests of the convergence hypothesis. *Journal of Econometrics*, **71**, 161-73
- Evans, P and G. Karras (1996). Convergence revisited. *Journal of Monetary Economics*, **27**, 249-65
- Hamilton, J.D. (1994), Time series analysis, Princeton, Princeton University Press.
- Harvey A.C. (1989), *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, Cambridge.
- Harvey, A.C. and J.Bernstein (2001). Measurement and testing of inequality from time series deciles with an application to US wages. *Review of Economics and Statistics* (to appear)
- Harvey, A.C., and S.J Koopman (1997). Multivariate structural time series models. In C. Heij *et al.*(eds). *System dynamics in economic and financial models*, 269-98: Chichester: Wiley and Sons.
- Hobijn, B. and P.H.Franses (2000), Asymptotically perfect and relative convergence of productivity. *Journal of Applied Econometrics*, 15, 59-81.
- Koopman S.J., Harvey, A.C., Doornik, J.A and Shephard, N. (2000). *STAMP 6.0 Structural Time Series Analyser, Modeller and Predictor*, London: Timberlake Consultants Ltd.
- Koopman, S.J. and Harvey, A.C. (2002) Computing observation weights for signal extraction and filtering, *Journal of Economic Dynamics and Control* (to appear)
- Kuo, B-S. and A. Mikkola (2001), How sure are we about purchasing power parity? Panel evidence with the null of stationary exchange rates. *Journal of Money Credit and Banking*

- Kwiatkowski, D., Phillips, P.C.B, Schmidt, P. and Y.Shin (1992), Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root ? *Journal of Econometrics* 44, 159-78.
- Nyblom, J., and A.C.Harvey (2000). Tests of common stochastic trends, *Econometric Theory*, 16, 176-99.
- O'Connell, P. (1998), The overvaluation of purchasing power parity, *Journal of International Economics*, 44, 1-19.
- Stock, J.H. (1991), Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Econometrics*, **28**, 435-59.