Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
International Journal of Forecasting 21 (2005) 667–686
international journal of forecasting
www.elsevier.com/locate/ijforecast

Growth, cycles and convergence in US regional time series

Vasco M. Carvalho[a], Andrew C. Harvey[b,*]

[a]Department of Economics, University of Chicago, 1126 E 59th Street, Chicago, IL 60637, USA
[b]Faculty of Economics, Sidgwick Avenue, Cambridge University, Cambridge CB3, 9DD, UK

Abstract

This article reports the results of fitting unobserved components (structural) time series models to data on real income per capita in eight regions of the United States. The aim is to establish stylised facts about cycles and convergence. It appears that while the cycles are highly correlated, the two richest regions have been diverging from the others in recent years. A new model is developed in order to characterise the converging behaviour of the six poorest regions. The model combines convergence components with a common trend and cycles. These convergence components are formulated as a second-order error correction mechanism which allows temporary divergence while imposing eventual convergence. After fitting the model, the implications for forecasting are examined. Finally, the use of unit root tests for testing convergence is critically assessed in the light of the stylised facts obtained from the fitted models.
© 2005 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

JEL classification: C32; O40
Keywords: Balanced growth; Error correction mechanism; Kalman filter; Signal extraction; Unobserved components

1. Introduction

The revival of growth theory in the mid-eighties has led to a substantial empirical research effort. The analysis of regional growth dynamics has proved to be particularly fertile ground for this literature: to the extent that common membership of a nation tends to assure factor mobility and to eliminate technological, preference and institutional differences, the basic assumptions of the neoclassical growth model are likely to be met, thus rendering regional data sets the ideal testbed for assessing the absolute convergence implications of the theory.

The approach adopted in this paper differs from most of the econometric literature in placing the emphasis on time series models for description and forecasting. We first show how fitting multivariate unobserved components (structural) time series models to data on real income per capita can help to establish stylised facts about cycles and convergence. Rather than simply using unit root tests to decide whether convergence is taking place, we explore different scenarios concerning the extent to which con-

* Corresponding author. Tel.: +44 1223 335228; fax: +44 1223 335475.
E-mail addresses: carvalho@uchicago.edu (V.M. Carvalho), Ach34@econ.cam.ac.uk (A.C. Harvey).

0169-2070/$ - see front matter © 2005 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2005.04.017

vergence is taking place and which regions – if any – are converging. A new unobserved components model is then developed and its dynamic properties are contrasted with those of a corresponding autoregressive formulation. Its principal feature is the introduction of what we call convergence components. These are combined with a common trend and then incorporated into a model with cycles and irregular components. The convergence components are formulated as a second-order error correction mechanism that allows temporary divergence. Fitting the model provides a description of the movements of the series in the past and enables coherent forecasts to be made.

In the light of what we find, it is apparent that tests – even when applied correctly – are at best of limited value, while at worst they are completely misleading. In any case, we believe a statistical description of what is happening coupled with a forecasting mechanism is of more value. This is the case in many areas of economics. For example, in the context of unit root testing and purchasing power parity (PPP), Maddala and Kim (1998, p. 138) state: 'A more important issue is not a test of the validity of PPP but an estimate of the time it takes for deviations from the PPP to correct themselves'.

1.1. Contradictory evidence on US regional convergence

A brief review of applied work on US regional growth serves not only to set the scene for our own empirical work but also to highlight the uncertainty in the literature. Looking at the results accumulated over the last two decades, one can find competing studies concluding in favour of absolute convergence, relative (conditional) convergence and divergence, depending on the approach taken.

Using cross-sectional data, Barro and Sala-i-Martin (1992) showed that a negative correlation between initial income per capita and growth was the norm for US regions. As in cross-country comparisons, they found a slow speed of convergence; see Sala-i-Martin (1996, p. 1326). Unlike in cross-country studies, this convergence was taken to be absolute: decreasing returns to scale should bring about a tendency for equalisation of income per capita across regions so that, in the long run, regions only display short term fluctuations around a common trend.

The validity of inferences drawn from the cross-sectional approach was questioned by Quah (1993) who showed that the 'beta convergence' of the cross-sectional studies resulted from a weighted average, and a negative value correlation between initial income per capita and growth meant only that the output differences between some pairs of countries had declined over the sample. It shed no light on heterogeneities or convergence clubs. Furthermore, Evans and Karras (1996) argued that the usual cross-sectional approach was only valid under incredible conditions. Durlauf and Quah (1999) provide a recent review of these issues.

These debates have led to an increasing interest in panel data methods for testing whether convergence is taking place and measuring the speed of convergence. Allowing for unobservable region-specific heterogeneity (individual effects), this literature has produced a very different picture of the regional convergence process in the US, characterised by rapid convergence to different steady states, that is relative convergence; see Evans and Karras (1996) and Evans (2000). However, these findings are problematic. Such a high speed of convergence is difficult to rationalise, even in traditional neoclassical growth models with a narrow view of capital. Econometrically, the typical dynamic panel formulation used is known to be subject to strong upward bias in short samples, as a result of correlation between lagged dependent variables and unobserved residuals; see Durlauf and Quah (1999). Moreover, in these short panels, fluctuations at the business cycle horizon and other high frequency movements introduce further bias, thus contributing to a systematic overestimation of the rate of convergence; again see Durlauf and Quah (1999).

Finally, there is the evidence from unit roots applied to univariate time series. Carlino and Mills (1993) find no convergence in time series of US regional per capita income, while Carlino and Mills (1993), Loewy and Papell (1996) and Tomljanovich and Vogelsang (2001) try to reconcile these results with the aforementioned evidence for convergence by allowing for trend breaks.

1.2. Outline

In Section 2, we review structural time series models (STMs), as implemented in the STAMP package of

Koopman, Harvey, Doornik, and Sheppard (2000), and show how they apply to series exhibiting balanced growth. Section 3 then employs these models to capture the stylised facts surrounding the movements in income per head in US regions. The slowly changing trends show the long-run movements from which we can infer any tendencies towards convergence. We differ from researchers such as Bernard and Durlauf (1996) in defining convergence in terms of the underlying trend rather than the observations.

Distinguishing trends from cyclical movements is essential to an effective study of convergence. However, the information in cycles is of considerable interest in itself. The recent paper by Carlino and Sill (2001) uses the methodology of Vahid and Engle (1993) to decompose the series on US regions into common trends and common cycles. We do not find the resulting cycles particularly plausible—for example they are rarely below zero. By contrast, our cycles, which are based on an UC model and calculated by a state space smoothing algorithm, are effectively based on two-sided filters rather than one-sided filters, and their movements are much closer to the NBER dating of expansions and recessions.

A multivariate model of growth and convergence is developed in Section 4 and applied in Section 5. The convergence mechanism is based on an error correction model. This can be incorporated into an unobserved components (UC) model that effectively decomposes trends into a common balanced growth path component and a set of convergence components. A key feature of the model is that the error correction mechanism is extended so as to produce smooth convergence components that can display temporary divergence, thereby rendering the notion of a simple measure of the speed of convergence open to question. However, definition 2 of Bernard and Durlauf (1996) is satisfied in that the forecast function for the difference between any pair of regions converges to a constant. Furthermore, because the cross-section is relatively small, proper account is taken of the cross-correlations across regions.

Section 6 investigates unit root tests for convergence. In doing so, we distinguish carefully between the notion of economies which have converged and those which are in the process of converging. We present the results of pairwise unit root tests and discuss the extent to which they are helpful in determining which regions can be grouped together in a convergence model. The conclusions are set out in Section 7.

2. Structural time series models and balanced growth

2.1. Univariate models

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

$$y_t = \mu_t + \varepsilon_t, \quad t = 1,..., T. \tag{1}$$

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

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim \mathrm{NID}\left(0, \sigma_\eta^2\right),$$
$$\beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim \mathrm{NID}\left(0, \sigma_\zeta^2\right), \tag{2}$$

where the irregular, level and slope disturbances, ε_t, η_t and ζ_t, respectively, are mutually independent and the notation NID(0,σ^2) denotes normally and independently distributed with mean zero and variance σ^2. If variances σ_η^2 and σ_ζ^2 are both 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, β. Allowing σ_ζ^2 to be positive, but setting σ_η^2 to zero gives an integrated random walk trend, which tends to be relatively smooth when estimated. The model is often referred is often referred to as the '*smooth trend*' model.

The statistical treatment of unobserved component models, as in the STAMP package of Koopman et al. (2000), 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, and maximizing it by an iterative procedure.

Distinguishing a long-term trend 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

$$y_t = \mu_t + \psi_t + \varepsilon_t, \quad t = 1,..., T. \tag{3}$$

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

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \quad t = 1,..., T, \tag{4}$$

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

$$\rho(\tau) = \rho^\tau \cos\lambda_c \tau, \quad \tau = 0, 1, 2,... \tag{5}$$

For $0 < \lambda_c < \pi$, the spectrum of ψ_t displays a peak, centred around λ_c, which becomes sharper as ρ moves closer to one; see Harvey (1989, p. 60). 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, that is setting σ_η^2 to zero, often allows a clearer separation into trend and cycle.

2.2. Convergence models

Long-run movements often have a tendency to converge to an equilibrium level. In an autoregressive framework, this is captured by an error correction model (ECM). The UC approach is to add cycle and irregular components to an ECM so as to avoid confounding the transitional dynamics of convergence with short-term steady-state dynamics. Thus,

$$y_t = \alpha + \mu_t + \psi_t + \varepsilon_t, \quad t = 1,..., T \tag{6}$$

with

$$\mu_t = \phi\mu_{t-1} + \eta_t \text{ or } \Delta\mu_t = (\phi - 1)\mu_{t-1} + \eta_t,$$
$$\eta_t \sim \text{NID}\left(0, \sigma_\eta^2\right).$$

Smoother transitional dynamics, and hence a better separation into convergence and short-term components, can be achieved by specifying μ_t in Eq. (6) as

$$\mu_t = \phi\mu_{t-1} + \beta_{t-1}, \quad t = 1,..., T,$$
$$\beta_t = \phi\beta_{t-1} + \zeta_t, \quad \zeta_t \sim \text{NID}\left(0, \sigma_\zeta^2\right) \tag{7}$$

when $\phi = 1$. *If, for convenience, we replace* β_{t-1} *by* β_t *in the equation for* μ_t, *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. The variance and ACF can be obtained from the properties of an AR(2) process or by noting that the model is a special case of the second-order cycle with $\lambda_c = 0$. The ℓ-step ahead forecast function, standardised by dividing by the current value of the gap, is $(1 + c\ell)\phi^\ell$, $\ell = 0,1,2,..$ where c is a constant that depends on the ratio, ω, of the gap in the current time period to the previous one, that is $\omega = \tilde{\mu}_{T|T}/\tilde{\mu}_{T-1|T}$. Since the one-step ahead forecast is $2\phi - \phi^2/\omega$, it follows that $c = 1 - \phi/\omega$, so

$$\tilde{\mu}_{T+\ell|T} = (1 + (1-\phi/\omega)\ell)\phi^\ell \tilde{\mu}_T, \quad \ell = 0, 1, 2,... \tag{8}$$

If $\omega = \phi$, the expected convergence path is the same as in the first order model. Convergence in the second-order model is typically much slower. Indeed, if the convergence process stalls sufficiently, the gap can be expected to widen in the short term. In the first-order model with $\phi = 0.9$, only 4% of the original gap is left after 30 time periods. By contrast, setting $\omega = 1$ in the second-order model leaves nearly 17% of the gap remaining.

2.3. Multivariate models

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

$$\boldsymbol{y}_t=\boldsymbol{\mu}_t+\boldsymbol{\psi}_t+\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t\sim \mathrm{NID}(\mathbf{0},\boldsymbol{\Sigma}_{\varepsilon}), \quad t=1,..,T, \tag{9}$$

where $\boldsymbol{\Sigma}_{\varepsilon}$ is an $N\times N$ positive semi-definite matrix. The trend is

$$\boldsymbol{\mu}_t=\boldsymbol{\mu}_{t-1}+\boldsymbol{\beta}_{t-1}+\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t\sim \mathrm{NID}\left(\mathbf{0},\boldsymbol{\Sigma}_{\eta}\right)$$

$$\boldsymbol{\beta}_t=\boldsymbol{\beta}_{t-1}+\boldsymbol{\zeta}_t, \quad \boldsymbol{\zeta}_t\sim \mathrm{NID}(\mathbf{0},\boldsymbol{\Sigma}_{\zeta}). \tag{10}$$

With $\boldsymbol{\Sigma}_{\eta}=\mathbf{0}$, we get the smooth trend model. With $\boldsymbol{\Sigma}_{\zeta}=\mathbf{0}$, we get the random walk plus drift.

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

$$\begin{bmatrix}\boldsymbol{\psi}_t\\ \boldsymbol{\psi}_t^*\end{bmatrix}=\left[\rho\begin{pmatrix}\cos\lambda_c & \sin\lambda_c\\ -\sin\lambda_c & \cos\lambda_c\end{pmatrix}\otimes\mathbf{I}_N\right]\begin{bmatrix}\boldsymbol{\psi}_{t-1}\\ \boldsymbol{\psi}_{t-1}^*\end{bmatrix}+\begin{bmatrix}\boldsymbol{\kappa}_t\\ \boldsymbol{\kappa}_t^*\end{bmatrix},$$
$$t=1,...T, \tag{11}$$

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

$$E\left(\boldsymbol{\kappa}_t\boldsymbol{\kappa}_t^{'}\right)=E\left(\boldsymbol{\kappa}_t^*\boldsymbol{\kappa}_t^{*'}\right)=\boldsymbol{\Sigma}_{\kappa}, \quad E\left(\boldsymbol{\kappa}_t\boldsymbol{\kappa}_t^{*'}\right)=0, \tag{12}$$

where $\boldsymbol{\Sigma}_{\kappa}$ 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.

2.4. Stability and balanced growth

The *balanced growth* UC model is a special case of Eq. (9):

$$\boldsymbol{y}_t=\boldsymbol{i}\mu_t+\boldsymbol{\alpha}+\boldsymbol{\psi}_t+\boldsymbol{\varepsilon}_t, \quad t=1,..,T, \tag{13}$$

where μ_t is a univariate stochastic trend as in Eq. (2), $\boldsymbol{i}$ 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 that 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 fixed or stochastic.

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 $\mathbf{D}\boldsymbol{i}=\mathbf{0}$, thereby rendering $\mathbf{D}\boldsymbol{y}_t$ 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. The multivariate stationarity test described in Nyblom and Harvey (2000) and Hobijn and Franses (2000) may be used to test the null hypothesis[1] of balanced growth; the test statistic is invariant to the choice of $\mathbf{D}$.

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

$$\Delta\boldsymbol{y}_t=\delta+\boldsymbol{\Gamma}\mathbf{D}\boldsymbol{y}_{t-1}+\sum_{r=1}^{p}\boldsymbol{\Phi}_r^*\Delta\boldsymbol{y}_{t-r}+\boldsymbol{\xi}_t,$$
$$\mathrm{Var}(\boldsymbol{\xi}_t)=\boldsymbol{\Sigma}_{\xi}, \tag{14}$$

where the $\boldsymbol{\Phi}_r^*$s are $N\times N$ matrices, $\mathbf{D}$ is as defined in the previous paragraph and the matrix $\boldsymbol{\Gamma}$ is $N\times(N-1)$. The system has a single unit root, guaranteed by the fact that $\mathbf{D}\boldsymbol{i}=\mathbf{0}$. 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 $\mathbf{D}\boldsymbol{y}_{t-1}$. For example, if $\mathbf{D}\boldsymbol{y}_t$ has elements $y_{it}-y_{i+1,t}$, $i=1,..,N-1$, 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, $\boldsymbol{\delta}=\beta(\mathbf{I}-\sum_{j=1}^{p}\boldsymbol{\Phi}_j^*)\boldsymbol{i}-\boldsymbol{\Gamma}\alpha$. Estimation by OLS applied to each equation in turn is fully efficient

[1] Hobijn and Franses (2000) actually say that they are testing whether the countries "*are converging*." In fact, stationarity tests are not useful if the countries are in the process of converging since initial values some way from the equilibrium will cause them to be rejected.

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 Eq. (14). This can be useful both 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.

2.5. Similar and common cycles

In the similar cycle model, the extent to which the cycles move together depends on the correlations between the disturbances driving them, since

$$\boldsymbol{\Sigma}_{\psi} = \left(1-\rho^2\right)^{-1}\boldsymbol{\Sigma}_{\kappa}.$$

Using principal components analysis, we can decompose $\boldsymbol{\Sigma}_{\kappa}$ as $\mathbf{EDE}'$ where $\mathbf{D}$ is a diagonal matrix of eigenvalues and $\mathbf{E}$ is the corresponding matrix of eigenvectors. The principal components themselves are contained in the series in the $N\times 1$ vector $\boldsymbol{\psi}_t^{\dagger}=\mathbf{E}'\boldsymbol{\psi}_t$, while the variance of the disturbances driving the j-th principal component is d_j, the j-th diagonal element of $\mathbf{D}$. These principal component cycles can be entered into model (9) by writing

$$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \mathbf{E}\boldsymbol{\psi}_t^{\dagger} + \boldsymbol{\varepsilon}_t. \tag{15}$$

The proportion of the variance of the i-th cycle accounted for by the j-th principal component is $e_{ij}^2 d_j/\sigma_{\psi i}^2$; note that $\sigma_{\psi i}^2=\sum_j e_{ij}^2 d_j$.

We can create a set of N standardised principal component cycles as $\boldsymbol{\psi}_t^{\ddagger}=\mathbf{D}^{-1/2}\mathbf{E}'\boldsymbol{\psi}_t$ with $\boldsymbol{\Sigma}_{\kappa}^{\ddagger}=\mathbf{I}$. The factor loadings are then $\boldsymbol{\Theta}=\mathbf{ED}^{1/2}$ and Eq. (15) becomes

$$\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Theta}\boldsymbol{\psi}_t^{\ddagger} + \boldsymbol{\varepsilon}_t.$$

This formulation is useful as a starting point for factor rotations.

If $\boldsymbol{\Sigma}_{\psi}$ is less than full rank, there are common cycles. A model can be estimated with a given rank as described in Koopman et al. (2000). If the rank of $\boldsymbol{\Sigma}_{\psi}$ is one, there is a single common cycle and the model can be written as

$$y_{it} = \mu_{it} + \theta_i\psi_t + \varepsilon_{it}, \quad i=1,\ldots,N \tag{16}$$

where ψ_t is a scalar cycle and the θ_is allow the common cycle to appear in each series with a different amplitude. One of the θ_is is set to be equal to unity and there is no need for a constant as in Eq. (13). A single common cycle is a common feature in the sense of Engle and Kozicki (1993) in that it may be removed by a linear combination, $\bar{\boldsymbol{\theta}}$, of the observations with the property that $\bar{\boldsymbol{\theta}}'\boldsymbol{\theta}=0$, where the $N\times 1$ vector $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_N)'$.

Testing the null hypothesis of a single common cycle is not straightforward. However, for the case of $N=2$, the distribution of the LR statistic is an even mixture of χ_0^2 and χ_1^2; see Harvey (1989, p. 236). Thus, the 5% critical value is 2.71.

3. Stylised facts: trends and cycles

In this section, we analyse trend and cyclical dynamics in the logarithms of real per capita incomes in US census regions: New England (NE), Mid East (ME), Great Lakes (GL), Plains (PL), South East (SE), South West (SW), Rocky Mountains (RM) and Far West (FW). The data were obtained from the Bureau of Economic Analysis and deflated by the US implicit price deflator (1996=100). Fig. 1 shows annual observations for the eight US census regions from 1950 to 1999. Carlino and Mills (1993) use annual data from 1929 (to 1990). However, because the fluctuations in the 1930s and 1940s are so much bigger than the cycles after 1950, it is difficult to model the whole series satisfactorily. The introduction of a trend break, as in Carlino and Mills (1993) and Loewy and Papell (1996), does not really address this problem. A corresponding analysis of quarterly seasonally adjusted observations for the same regions from 1969:1 to 1999:4 can be found in Carvalho and Harvey (2002). The conclusions are similar.

We report the results of fitting unrestricted multivariate structural time series models of the form (9) with smooth trends in order to obtain some idea of stylised facts. Estimation of this and all other models below was done using program routines written in the OX 3.0 language (Doornik, 1999) with use being made of the SsfPack package for state space algorithms of Koopman, Shephard and Doornik (1999).

All parameters were estimated by maximum likelihood as described in Section 2.1 and variances are

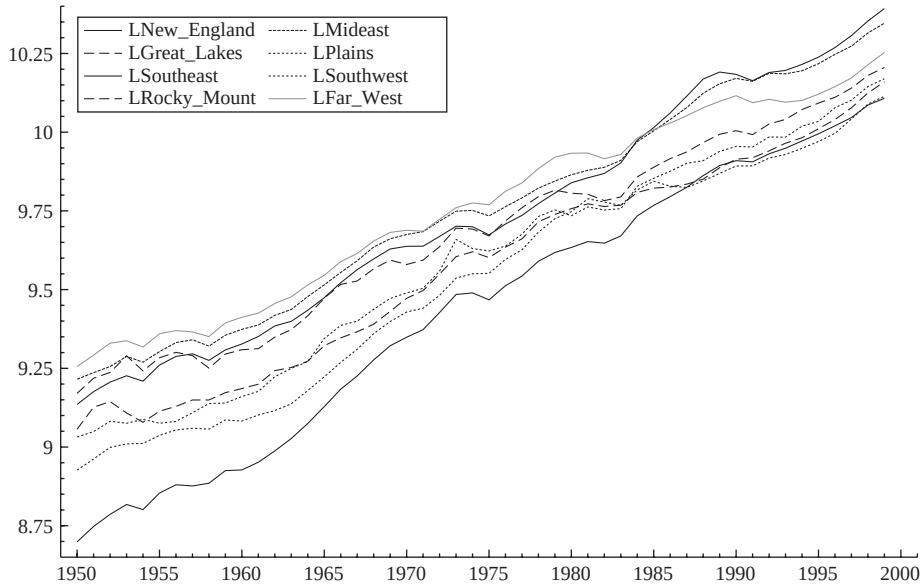

Fig. 1. Annual income per capita in eight US census regions.

reported multiplied by 10^7. The estimated covariance matrices are reported by showing the variances on the main diagonal while the entries above contain the cross-correlations. All graphs show the estimated components as extracted by the state-space smoothing algorithm.

3.1. Cycles

The smoothed cyclical components, $\tilde{\psi}_{it|T}$, for the eight annual regional series are shown in Fig. 2. The recessions of 1954, 1961, 1970, 1975, 1980, 1982 and 1991 all show up with a high degree of coherence

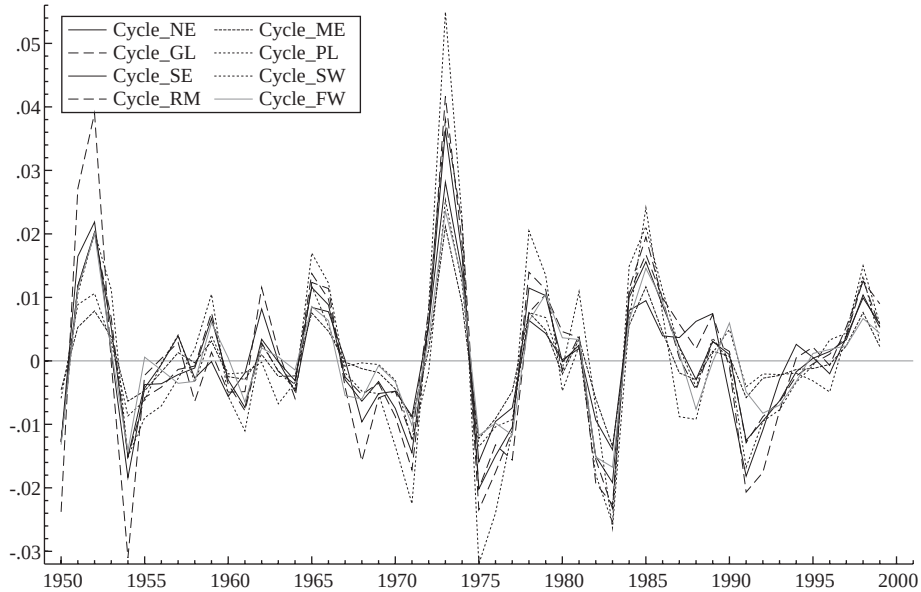

Fig. 2. Cyclical components for annual data.

across regions, unlike in Carlino and Sill (2001). The period of the cycle is 5.3 years with a damping factor, ρ, of 0.80. The period for the quarterly series is 4.25 years with a damping factor of 0.85.

The matrix of variances and cross-correlations obtained from $\tilde{\boldsymbol{\Sigma}}_\kappa$ is shown below. There are considerable differences in volatility, with the variance of the disturbances in PL being almost six times as great as that of ME. These findings are similar to those reported by Carlino and Sill (2001, p. 452). However, our ordering of the regions in terms of volatility differs from theirs. Furthermore, we find that the richest regions (NE, ME and FW) are those with the least volatile cyclical components.

$$\begin{bmatrix} 376.9 & 0.863 & 0.913 & 0.882 & 0.932 & 0.735 & 0.962 & 0.851 \\ & 172.2 & 0.975 & 0.966 & 0.948 & 0.927 & 0.818 & 0.848 \\ & & 646.8 & 0.987 & 0.994 & 0.865 & 0.893 & 0.921 \\ & & & 1014.1 & 0.976 & 0.816 & 0.837 & 0.864 \\ & & & & 508.8 & 0.822 & 0.924 & 0.942 \\ & & & & & 420.8 & 0.720 & 0.790 \\ & & & & & & 876.4 & 0.932 \\ & & & & & & & 332.7 \end{bmatrix} \begin{matrix} \text{NE} \\ \text{ME} \\ \text{GL} \\ \text{PL} \\ \text{SE} \\ \text{SW} \\ \text{RM} \\ \text{FW.} \end{matrix}$$

There are high, positive correlations between the cyclical disturbances in all regions. The first principal component of $\tilde{\boldsymbol{\Sigma}}_\kappa$ accounts for 91% of the total variance while the second accounts for a further 5%. Table 1 shows the weights applied to each of the first two principal component cycles to obtain the regional cycle and the percentage variance in each regional cycle contributed by the first two components; see Eq. (15) and the text below it. The e_{ij}'s not only show the weight given to each principal component in forming the cycle for each region, but they are also the weights given to the regional cycles in constructing the principal components. The first principal component is a composite of the regional cycles in which the weights are all positive and of a similar order of magnitude. The second principal component contrasts the regional cycles.

Table 1
Weights assigned to the first two principal components in cycles for each series

Region	Weight, e_{i1}	Weight, e_{i2}	Proportion, 1	Proportion, 2
NE	0.29	0.29	0.89	0.05
ME	0.20	−0.23	0.92	0.07
GL	0.40	−0.18	0.99	0.01
PL	0.49	−0.38	0.94	0.03
SE	0.36	−0.01	0.99	0.0003
SW	0.28	−0.45	0.73	0.10
RM	0.44	0.68	0.88	0.11
FW	0.27	0.17	0.88	0.02

By standardising the weights, e_{i1}, for the first principal component so that they sum to one, it is possible to construct a first principal component cycle[2] from the individual regional cycles. For annual data, this principal component is shown in Fig. 3, where it is contrasted with the smoothed cyclical component series extracted from the univariate US annual series. The two cycles are very close. This again illustrates the point that the main source of regional cyclical volatility is a component closely related to the national business cycle.

Since the first principal component is so dominant, we decided to see what happens when the model is estimated with a single common cycle. The results were not particularly useful, in that the cycle obtained bore no relation to the cycles shown in Fig. 3. This is perhaps an indication that the common cycle restriction is too strong.

The other stationary component is the irregular component, the variances and cross-correlations of which are shown below. It is dominated by the cycle (remember that the disturbance variance needs to be divided by $1-\rho^2$ to give the cycle variance), but nevertheless plays a useful role. There are no general conclusions to be drawn from the pattern of correlations.

$$\begin{bmatrix} 286.2 & 0.991 & 0.877 & -0.581 & 0.824 & 0.621 & 0.425 & 0.941 \\ & 452.1 & 0.881 & -0.477 & 0.775 & 0.586 & 0.461 & 0.929 \\ & & 980.9 & -0.362 & 0.768 & 0.442 & 0.205 & 0.737 \\ & & & 849 & -0.536 & -0.317 & -0.340 & -0.517 \\ & & & & 316.15 & 0.738 & 0.018 & 0.806 \\ & & & & & 196.8 & -0.269 & 0.826 \\ & & & & & & 41.3 & 0.259 \\ & & & & & & & 402.4 \end{bmatrix} \begin{matrix} \text{NE} \\ \text{ME} \\ \text{GL} \\ \text{PL} \\ \text{SE} \\ \text{SW} \\ \text{RM} \\ \text{FW} \end{matrix}$$

3.2. Trends

The smooth trends, $\tilde{\mu}_{it|T}$, extracted from the annual series by the state space smoothing algorithm are shown in Fig. 4. A first glance seems to indicate that cross-sectional dispersion has declined, thus indicating convergence. However, closer inspection reveals otherwise. In particular, note how the trend dynamics of NE and ME, the two wealthiest regions at the end of the sample, differ from the

[2] Specifically, this is the first element of $\mathbf{E}'\tilde{\psi}_{t|T}$ and so the (unstandardised) weights are as in the first column of $\mathbf{E}$ and it is this that is reproduced under the heading e_{i1}.

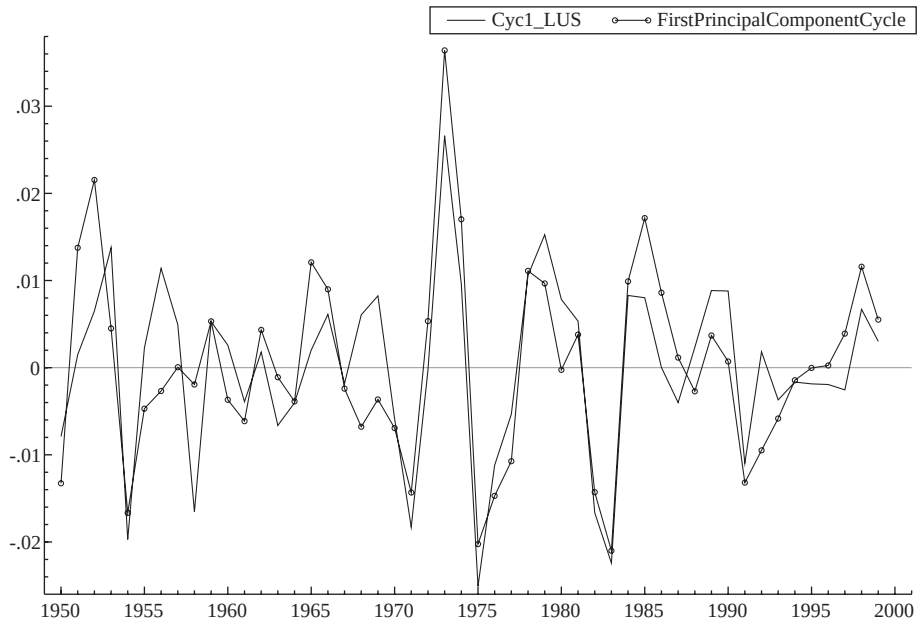

Fig. 3. Cycle given by first principal component and US aggregate cycle.

remaining six regions. Thus, while other relatively rich regions in the 1950s such as the Far West and Great Lakes seem to be converging to the levels of lower income regions, New England and Mid East, particularly from the late 1970s onwards, show no clear tendency to converge to some kind of common national trend.

The above conclusions are confirmed by the plot of the cross-sectional standard deviation of the smoothed trend component, SD($\tilde{\mu}_{it|T}$), in Fig. 5. Overall cross-

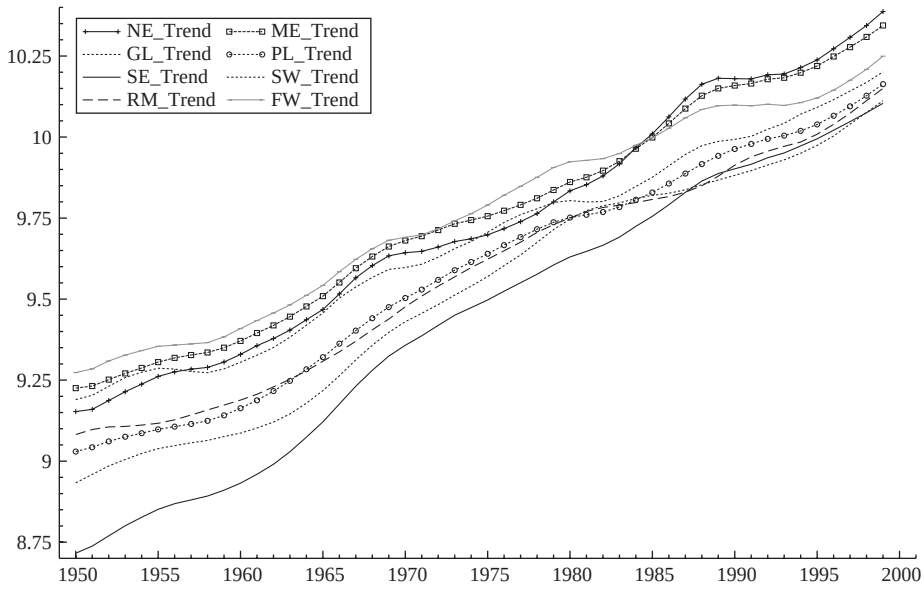

Fig. 4. Smooth trends for annual data.

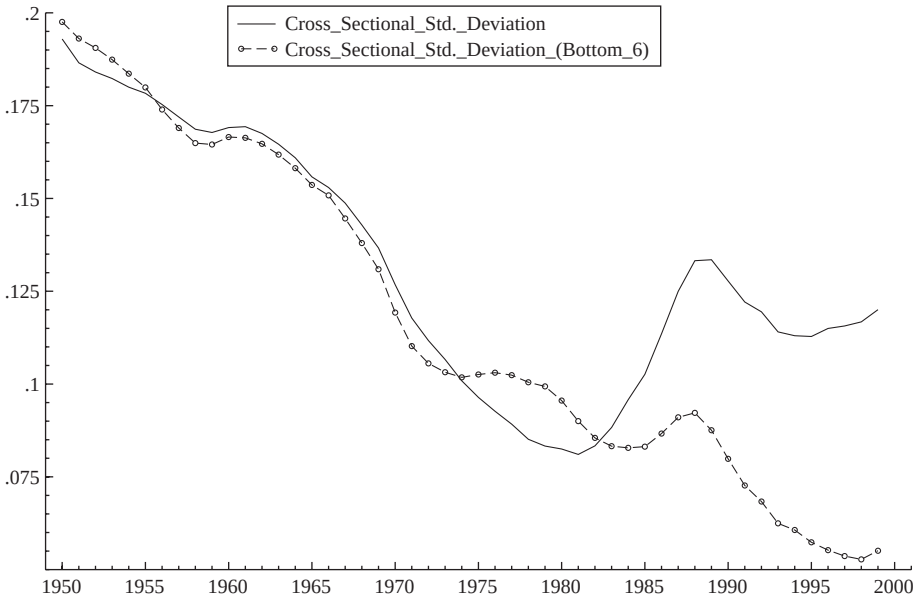

Fig. 5. Cross-sectional standard deviation.

sectional dispersion, which was clearly declining until the early eighties, has since reversed its tendency so that, by the end of the sample, the situation is much the same as in the late sixties. In other words, across all regions, divergence rather than convergence seems to have been the rule since the early eighties. Excluding the two richest regions alters this conclusion. For the remaining six regions (the poorest as defined at the end of the sample), the fall in trend dispersion seems to have been continuing throughout the last two decades. Hence, there appears to be convergence of all regions apart from the two richest which are diverging from the other six regions as well as from each other.

The estimated variances and cross-correlations of the slope disturbances are reproduced below. As in the cyclical component, considerable regional differences are evident in trend volatility. For example, the variance of the trend component in NE is almost three times as great as that of SW. This is reflected in the larger fluctuations observed in the smoothed NE trend.

$$\begin{bmatrix} 1310.1 & 0.969 & 0.717 & 0.640 & 0.840 & 0.414 & -0.154 & 0.849 \\ & 665.2 & 0.716 & 0.771 & 0.891 & 0.422 & 0.005 & 0.859 \\ & & 1183.7 & 0.725 & 0.822 & 0.352 & -0.189 & 0.816 \\ & & & 412.8 & 0.819 & 0.569 & 0.473 & 0.861 \\ & & & & 546.5 & 0.603 & 0.068 & 0.799 \\ & & & & & 579.5 & 0.554 & 0.557 \\ & & & & & & 437.8 & 0.162 \\ & & & & & & & 935.6 \end{bmatrix} \begin{matrix} \text{NE} \\ \text{ME} \\ \text{GL} \\ \text{PL} \\ \text{SE} \\ \text{SW} \\ \text{RM} \\ \text{FW} \end{matrix}$$

The cross-correlations between the trend disturbances show considerably more variation than those for the cycles. While the disturbances in the two wealthiest regions are highly correlated, some of the other disturbances actually display a negative correlation.

4. Multivariate convergence models

The multivariate STM has provided an idea of possible convergence patterns. We now proceed to develop a model to capture these movements.

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

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

with

$$\mu_{it} = \sum_{j=1}^{N} \phi_{ij}\mu_{j,t-1} + \eta_{it}, \quad i = 1,...,N \tag{17}$$

with $\sum_{j=1}^{N}\phi_{ij}=1$ for $i=1,...,N$. 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

$$\Delta\mu_{it} = \sum_{j\neq i} \phi_{ij}\left(\mu_{j,t-1} - \mu_{i,t-1}\right) + \eta_{it}, \quad i = 1,\ldots,N, \tag{18}$$

shows how the growth of the i-th economy depends on the gap between it and each of the other $N-1$ economies. In the bivariate case, the model implies that the difference between the two economies, $y_t = y_{1t} - y_{2t}$, satisfies

$$\Delta y_t = (\phi - 1)(y_{t-1} - \alpha) + \eta_t = \delta + (\phi - 1)y_{t-1} + \eta_t, \quad t = 2,\ldots,T, \tag{19}$$

where $\eta_t = \eta_{1t} - \eta_{2t}$, $\alpha = \alpha_1 - \alpha_2$, $\phi = 1 - \phi_{12} - \phi_{21}$ and $\delta = \alpha(1-\phi)$. 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. Writing the model in this 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).

In matrix form

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

with

$$\boldsymbol{\mu}_t = \boldsymbol{\Phi}\boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \quad \mathrm{Var}(\boldsymbol{\eta}_t) = \boldsymbol{\Sigma}_\eta \tag{20}$$

or, in error correction form,

$$\Delta\boldsymbol{\mu}_t = (\boldsymbol{\Phi} - \mathbf{I})\boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, \quad \mathrm{Var}(\boldsymbol{\eta}_t) = \boldsymbol{\Sigma}_\eta. \tag{21}$$

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

If we write

$$\bar{\boldsymbol{\phi}}'\Delta\boldsymbol{\mu}_t = \bar{\phi}'(\boldsymbol{\Phi} - \mathbf{I})\boldsymbol{\mu}_{t-1} + \bar{\boldsymbol{\phi}}'\boldsymbol{\eta}_t \tag{22}$$

it is clear that the $N \times 1$ vector of weights, $\bar{\phi}$, which gives a random walk, must be such that $\bar{\phi}'(\boldsymbol{\Phi} - \mathbf{I}) = \mathbf{0}'$. Since the roots of $\boldsymbol{\Phi}'$ are the same as those of $\boldsymbol{\Phi}$, it follows from writing $(\boldsymbol{\Phi}' - \mathbf{I})\bar{\boldsymbol{\phi}} = \mathbf{0}$ that $\bar{\boldsymbol{\phi}}$ is the eigenvector of $\boldsymbol{\Phi}'$ 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.[3] The inclusion of a time trend in the model means that the overall common trend is a random walk with drift, β, and if $\boldsymbol{\alpha}$ is defined such that $\boldsymbol{\alpha}'\bar{\boldsymbol{\phi}} = 0$, each element of $\boldsymbol{\alpha}$ is a deviation from the common trend.

Unobserved components models. The model in Eq. (17) may be extended so as to include cycle and irregular components. Thus

$$y_{it} = \alpha_i + \beta t + \mu_{it} + \psi_{it} + \varepsilon_{it}, \quad i = 1,\ldots,N. \tag{23}$$

In matrix terms, using the notation of Eq. (9),

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \beta \boldsymbol{i} t + \boldsymbol{\mu}_t + \boldsymbol{\psi}_t + \boldsymbol{\varepsilon}_t, \tag{24}$$

with $\boldsymbol{\mu}_t$ as in Eq. (20). The model with a smooth convergence mechanism is written as:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\alpha} + \boldsymbol{\mu}_t + \boldsymbol{\psi}_t + \boldsymbol{\varepsilon}_t, \\ \boldsymbol{\mu}_t &= \boldsymbol{\Phi}\boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}_{t-1} \\ \boldsymbol{\beta}_t &= \boldsymbol{\Phi}\boldsymbol{\beta}_{t-1} + \boldsymbol{\zeta}_t, \end{aligned} \tag{25}$$

so that $\boldsymbol{\mu}_t$ is driven by an $N \times 1$ vector of slopes, $\boldsymbol{\beta}_t$, that evolve over time because of a disturbance vector, $\boldsymbol{\zeta}_t$, with covariance matrix, $\boldsymbol{\Sigma}_\zeta$. When $\boldsymbol{\Phi} = \mathbf{I}$, the $\boldsymbol{\mu}_t$ vector reduces to a set of smooth trend components, as in the model applied in the previous section. As with the first-order model, Eq. (20), the forecasts from the second-order model converge to paths parallel to that of the common trend, $\bar{\mu}_{\phi t}$. Thus, the forecast convergence condition of Bernard and Durlauf (1996, definition 2) is satisfied.

VECM. Convergence may be captured by the common trend VECM of Eq. (14). The matrix $\boldsymbol{\Gamma}$ contains $N(N-1)$ free parameters and these may be estimated by OLS applied to each equation in turn. The $\boldsymbol{\Phi}$ matrix of Eq. (17) may then be estimated[4] as it is

[3] This is because $\lim_{j\to\infty}\boldsymbol{\Phi}^j = \boldsymbol{i}\bar{\boldsymbol{\phi}}'$; 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, p. 681).

[4] We can also directly adopt the parameterisation implicit in the $\boldsymbol{\Phi}$ 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.

given by $\mathbf{\Gamma D}+\mathbf{I}$. However, there is no guarantee that the estimate of $\mathbf{\Gamma}$ will be such that $N-1$ of the roots of $\mathbf{\Phi}$ have modulus less than one. If the vector of gaps, $\boldsymbol{\alpha}$, is parameterised with respect to the contrasts in $\mathbf{D}\mathbf{y}_{t-1}$, then $\boldsymbol{\delta}=\beta(\mathbf{I}-\sum_{j=1}^{p}\mathbf{\Phi}_j^*)\boldsymbol{i}-\mathbf{\Gamma}\boldsymbol{\alpha}$.

4.1. Deviation and benchmark restrictions

As it stands, the UC model is difficult to estimate because $\mathbf{\Phi}$ contains $N(N-1)$ parameters. One way 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) of 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

$$\bar{\mu}_{w,t}=\sum_{i=1}^{N}w_i\mu_{it},\qquad \sum_{i=1}^{N}w_i=1$$

and set

$$\begin{aligned}&\phi_{ii}=\pi_i+1-\pi_i w_i=\pi_i(1-w_i)+1\quad\text{and}\\&\phi_{ij}=-\pi_i w_j,\quad i\neq j.\end{aligned}\tag{26}$$

Substituting in Eq. (18) yields

$$\begin{aligned}\Delta\mu_{it}&=\pi_i\sum_{j\neq i}w_j\left(\mu_{i,t-1}-\mu_{j,t-1}\right)+\eta_{it}\\&=\pi_i\left(\mu_{i,t-1}-\bar{\mu}_{w,t-1}\right)+\eta_{it},\quad i=1,..,N.\end{aligned}\tag{27}$$

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.

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

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

$$\begin{aligned}&\Delta\left(\mu_{it}-\bar{\mu}_{\phi,t}\right)=\pi\left(\mu_{i,t-1}-\bar{\mu}_{\phi,t-1}\right)+\eta_{it}-\bar{\eta}_{\phi,t},\\&i=1,\ldots,N,\end{aligned}\tag{28}$$

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 condition[5] is $-2<\pi<0$.

There are a number of ways to proceed. If both π_is and w_js 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_is or the w_js. If we let the π_is be the same, we can estimate the w_js 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_js and estimate the N π_is. For the case $N=2$, these two options are equivalent. When the π_is 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 is, and let the corresponding π_is be free. If n is set to zero, we then, for $n<N-1$, have $N-n-1$ free w_is to estimate, together with n π_is 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_is since, provided $\pi_j<0$ for all j, $\bar{\phi}_i=(1/\pi_i)/\sum_j(1/\pi_j)$ as is easily seen[6] from Eq. (27). If we regard it as reasonable to have $0\leq\bar{\phi}_i\leq 1$, then the π_is must be less than or equal to zero. If a $\pi_i=0$, then we get a benchmark model, as $\bar{\phi}_i\rightarrow 1$ as $\pi_i\rightarrow 0$. Within the context of Eq. (27), a test of $\pi_i=0$ can

[5] The matrix $\mathbf{i}\mathbf{w}'$ 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 $\mathbf{\Phi}=(1+\pi)\mathbf{I}-\pi\mathbf{i}\mathbf{w}'$ also has a single unit root while the rest are $1+\pi$.

[6] In matrix terms, $\mathbf{\Phi}=\mathbf{I}+\mathbf{\Pi}_D-\boldsymbol{\pi}\boldsymbol{i}'$, where $\mathbf{\Pi}_D$ is a diagonal matrix with the elements of $\boldsymbol{\pi}$ on its diagonal. We want to find ϕ such that $\mathbf{\Phi}'\bar{\boldsymbol{\phi}}=(\mathbf{I}+\mathbf{\Pi}_D-\boldsymbol{i}\boldsymbol{\pi}')\bar{\boldsymbol{\phi}}=\bar{\boldsymbol{\phi}}$. 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.

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

$$\Delta\mu_{it} = \pi_i\left(\mu_{i,t-1} - \mu_{N,t-1}\right) + \eta_{it}, \quad i = 1,\ldots,N-1,$$

$$\Delta\mu_{Nt} = \eta_{Nt},$$

where the roots of the transition matrix are one and π_i+1, $i=1,..,N-1$ so $-2<\pi_i<0$, $i=1,..,N-1$ for convergence. Note that this model is a special case of Eq. (27) obtained by setting all the weights apart from w_N equal to zero. Since μ_{Nt} is a random walk, we have

$$\Delta(\mu_{it} - \mu_{Nt}) = \pi_i\left(\mu_{i,t-1} - \mu_{N,t-1}\right) + \eta_{it} - \eta_{Nt}, \quad i = 1,\ldots,N-1. \tag{29}$$

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 logarithm of the weighted sum of the original observations. Working in logarithms has no implications for the benchmark model.

4.2. Autoregressive models

The deviation and benchmark constraints can be incorporated into an autoregressive model because $y_{i,t-1}-\bar{y}_{w,t-1}$, $i=1,..,N$ are all co-integrating vectors. If one is dropped, it can be reconstructed as a linear combination of the others. Thus, the $\boldsymbol{\Gamma}$ and $\mathbf{D}$ matrices in Eq. (14) can be formed with suitable constraints. However, it is more convenient to set up the model as

$$\Delta y_{it} = \delta_i + \pi_i\left(y_{i,t-1} - \bar{y}_{w,t-1}\right) + \sum_{j=1}^{N}\sum_{r=1}^{p}\phi^*_{ijr}\Delta y_{j,t-r} + \eta_{it}, \quad i = 1,..,N. \tag{30}$$

The parameters may be efficiently[7] 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_is, we can solve to get β and a set of α_is for relative convergence since

$$\delta_i = \beta\left[1 - \sum_{j=1}^{N}\sum_{r=1}^{p}\phi^*_{ij}\right] - \pi_i\alpha_i, \quad i = 1,..,N. \tag{31}$$

In a benchmark model with $\pi_N=0$, there are $N-1$ gaps represented by α_i, $i=1,\ldots,N-1$. 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_is are negative, so that the equation $\sum_i\alpha_i/\pi_i=0$ can be added to those in Eq. (31).

As N becomes large, the above AR model runs into difficulties because of the potentially large number of ϕ^*_{ijr} parameters, N^2p in all. It may well be the case that little explanatory power is lost by only including lagged differences of y_{it}, and it would be interesting to explore the implications of this for forecasting.

4.3. Unobserved components model

In the autoregressive framework, the natural way to proceed when the restrictions in Eq. (26) are imposed is to estimate π_is for a given set of preassigned w_is. An unobserved components formulation, however, requires nonlinear optimisation with respect to the elements of $\boldsymbol{\Phi}$ (as well as the other parameters, such as variances of disturbances). Since it is unclear what constraints should be imposed on the π_is, it is relatively more attractive to assume a *homogeneous* model in which $\pi_i=\pi$, $i=1,\ldots,N$, and to estimate the $\bar{\phi}_i$ parameters constraining them to lie between zero and one and to sum to one; these constraints can be imposed by employing a logistic transformation and maximizing with respect to $N-1$ unconstrained parameters, ξ_i, $i=1,..,N-1$, defined by $\bar{\phi}_i=\bar{\phi}^*_i\exp(\xi_i)/(1+\exp(\xi_i))$, where $\bar{\phi}^*_i=1-\sum_{j=1}^{i-1}\bar{\phi}_j$, $j=2,\ldots,N-1$ and $\bar{\phi}^*_1=1$. It is further assumed that the convergence parameter, $\phi=1+\pi$, is such that $0\le\phi\le 1$, with $\phi=1$ indicating no convergence. The statistical treatment of the model is based on the state space form, with the $\boldsymbol{\mu}_t$ vector initialised with a diffuse prior.

[7] If the general model, Eq. (14), can be estimated, an LR test of the constraints implied by Eq. (30) can be carried out.

The model can be rewritten so as to consist of convergence components, $\mu_{it}^{\dagger}$, $i=1,\ldots,N$, which are deviations from the common trend, $\bar{\mu}_{\phi,t}$. Then

$$y_{it} = \alpha_i + \bar{\mu}_{\phi,t} + \mu_{it}^{\dagger} + \psi_{it} + \varepsilon_{it}, \quad i = 1,..,N, \qquad (32)$$

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

$$\mu_{it}^{\dagger} = \phi\mu_{i,t-1}^{\dagger} + \eta_{it}^{\dagger}, \quad i = 1,\ldots,N-1, \qquad |\phi|<1, \quad (33)$$

and

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

If the state vector is defined in terms of the common trend and convergence components only $N-1$ of the latter need be included as $\sum_{i=1}^{N}\bar{\phi}_i\mu_{it}^{\dagger}=0$.

The extension to the smooth convergence processes is straightforward: Eq. (33) is replaced by

$$\mu_{it}^{\dagger} = \phi\mu_{i,t-1}^{\dagger} + \beta_{i,t-1}^{\dagger}, \quad \beta_{it}^{\dagger} = \phi\beta_{i,t-1}^{\dagger} + \zeta_{it}^{\dagger},$$
$$i = 1,\ldots,N-1, \qquad (34)$$

with $0\le\phi\le1$, while $\bar{\mu}_{\phi,t}$ is an integrated random walk.

If, say, $\bar{\phi}_N=1$, then the N-th series is a benchmark and $\bar{\mu}_{\phi t}$ is replaced by μ_{Nt}; in this case, we can have different ϕ_is in Eqs. (33) and (34).

Once the series have converged, the $\mu_{it}^{\dagger}$s become stationary with variances, for Eq. (34), given by $\sigma_i^{\dagger 2}(1+\phi^2)/(1-\phi^3)^2$, where $\sigma_i^{\dagger 2}$ is the i-th diagonal element of the covariance matrix of the vector $\boldsymbol{\zeta}_t^{\dagger}=(\zeta_{1t}^{\dagger},\zeta_{2t}^{\dagger},..,\zeta_{Nt}^{\dagger})'=\boldsymbol{\zeta}_t-\boldsymbol{i}\bar{\zeta}_{\phi,t}$. Since $\bar{\zeta}_{\phi,t}=\bar{\boldsymbol{\phi}}'\boldsymbol{\zeta}_t$, this covariance matrix is $\mathrm{Var}(\boldsymbol{\zeta}_t^{\dagger})=(\boldsymbol{\zeta}_t-\boldsymbol{i}\bar{\boldsymbol{\phi}}')'\,\boldsymbol{\Sigma}_{\zeta}(\boldsymbol{\zeta}_t-\boldsymbol{i}\bar{\boldsymbol{\phi}}')$.

5. Convergence and divergence in US regions

The preliminary investigation of stylised facts reported in Section 4 indicates that the two richest regions, NE and ME, follow growth paths which, especially for the last two decades, seem to be diverging from the growth paths of the other regions. Hence, we only fit a convergence model to the six poorer regions. The results reported are for the homogeneous model, Eq. (32), with smooth convergence as in Eq. (34) and absolute convergence, that is $\alpha_i=0$, $i=1,\ldots,N$.

The similar cycle damping factor was estimated to be 0.79 while the period was 8.0 years. The estimated variances and cross-correlations of $\boldsymbol{\Sigma}_{\kappa}$ are

1795.3	0.867	0.976	0.887	0.553	0.914	GL
	1373.7	0.938	0.763	0.854	0.842	PL
		1163.4	0.914	0.720	0.956	SE
			1166.5	0.662	0.990	SW
				1296.5	0.721	RM
					872.8	FW

while for the irregular component

834.5	−0.357	0.716	0.534	0.810	0.675	GL
	774.7	−0.706	−0.518	−0.433	−0.515	PL
		215.1	0.773	0.787	0.821	SE
			0.5	0.918	0.984	SW
				138.9	0.972	RM
					308.1	FW

The convergence parameter, ϕ, was estimated as 0.889, the estimates of the common trend weights, $\bar{\phi}_i$, are

$\bar{\Phi}_{GL}$	$\bar{\Phi}_{PL}$	$\bar{\Phi}_{SE}$	$\bar{\Phi}_{SW}$	$\bar{\Phi}_{RM}$	$\bar{\Phi}_{FW}$
0.640	0.298	0.006	0.0004	0.056	0.00003

and the estimated variances and cross-correlations of $\boldsymbol{\Sigma}_{\zeta}$ are

354.9	0.946	0.573	−0.463	0.074	0.809	GL
	205.2	0.696	−0.170	0.269	0.759	PL
		246.3	−0.052	−0.071	0.399	SE
			237.5	0.552	−0.270	SW
				92.9	−0.193	RM
					576.3	FW

Recall that this specification not only allows us to separate trends and cycles but also separates out the long-run balanced growth path from the transitional regional dynamics, thus permitting a characterisation of convergence stylised facts. We will focus on this feature of the model, since the cycle is of secondary interest here and there is little information beyond what was presented in Section 4. Fig. 6 shows the smoothed estimates of the convergence components, $\mu_{it}^{\dagger}$, for the six regions, while Fig. 7 displays the estimated common trend, $\bar{\mu}_{\phi,t}$, together with the estimated trends for each region. The plot of the cross-sectional standard deviation computed from the smoothed trends is similar to that shown in Fig. 5.

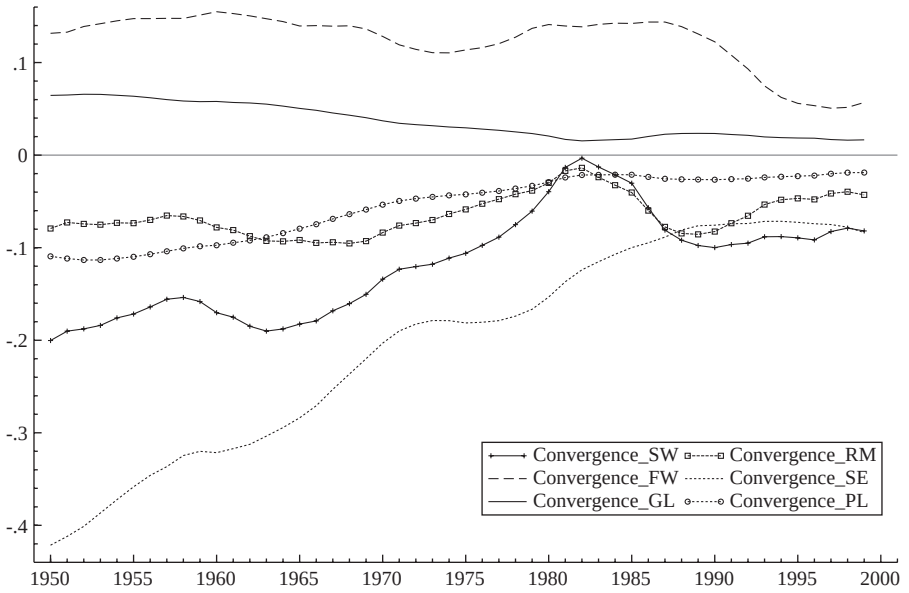

Fig. 6. Convergence component for six poorest regions.

The convergence process is such that the common trend is essentially constructed by weighting Great Lakes two-thirds and Plains one third. This suggests that we could proceed further by constructing a benchmark from the Great Lakes and Plains, thereby allowing the assumption of a single rate of convergence to be relaxed[8]; see the discussion in the middle of Section 4.1.

Although the convergence is clear, substantial heterogeneity is evident with the convergence components being neither monotonic over time nor homogeneous across regions. Thus, while Great Lakes and Plains display smooth dynamics with most of the convergence towards the common trend taking place from 1950 to 1980, Rocky Mountains and South West display stronger convergence dynamics but only from the mid sixties onwards. Moreover, for the latter group, the 1980s are actually a period of strong divergence that is only reversed in the last decade of the sample. This type of process is also evident in the Far West region where, following a period of (slow) convergence up until the early seventies, diverging dynamics dominate from the mid seventies up until the late eighties when convergence (now stronger) resumes. Finally, the high average rate of growth of the South East region translates into the strong catching-up process displayed up until the late eighties. However, during the nineties, this process appears to have slowed down, or even reversed.

The estimation of the convergence model leads to slight changes in the analysis of trend variability and cross-regional correlations. The trend volatility for all regions, indicated by $\tilde{\boldsymbol{\Sigma}}_{\zeta}$, is now lower than in the model fitted in Section 4, while there are more negative cross-correlations.

Multi-step forecasts enable us to analyse further the convergence dynamics implied by the estimates. Fig. 8 shows the forecasts of the convergence components for the six regional series over a 20-year horizon (2000–2019), while Fig. 9 contrasts the different paths. The striking feature of Figs. 8 and 9 is not the eventual convergence, but rather the prediction of divergence in the short run. Thus, although Plains and Great Lakes converge rapidly to the growth path of the common trend, which is hardly surprising given the composition of the common trend, the Far West, Rocky Mountains, South East and South West are all expected to widen their income gap,

[8] To be more specific we could set $\bar{\phi}_i$ to zero for all regions – apart from GL and PL – and let the corresponding ϕ_i be specific to each region. GL and PL have the same ϕ_i.

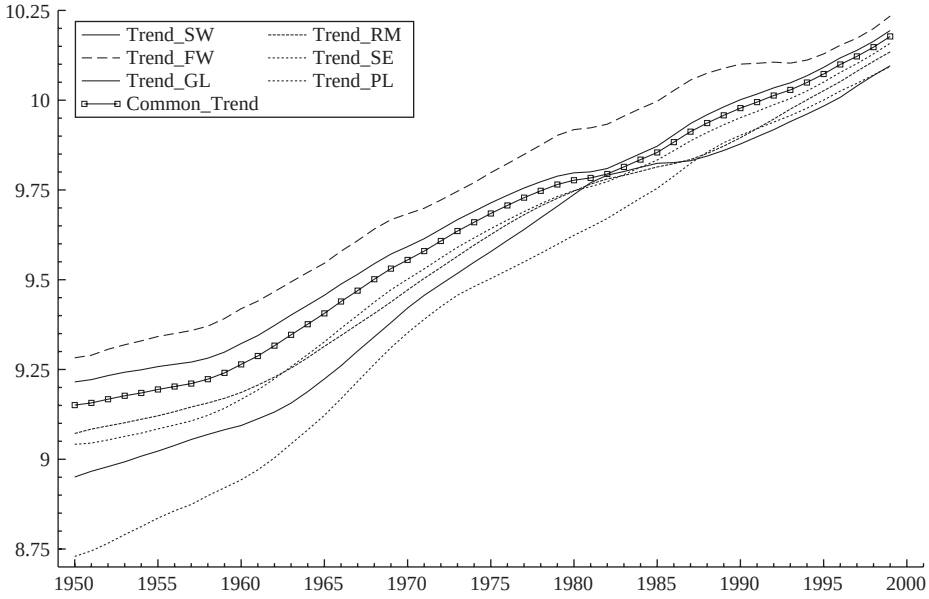

Fig. 7. Common trend and growth paths for six poorest regions.

relative to the common trend, during the first 5 years of the forecast period. Only then do they resume their convergence towards the common trend and even then with noticeable differences in dynamics. Thus, by 2019 (2009), while the Great Lakes are expected to have removed 95% (50%) of the 1999 gap, the remaining six regions will have removed only between 51% (17%) – for the SW – and 60% (28%) – for RM – of their respective gaps. This temporary divergence is a feature of the smooth convergence model. The second-order error correction specification not only allows slower changes but

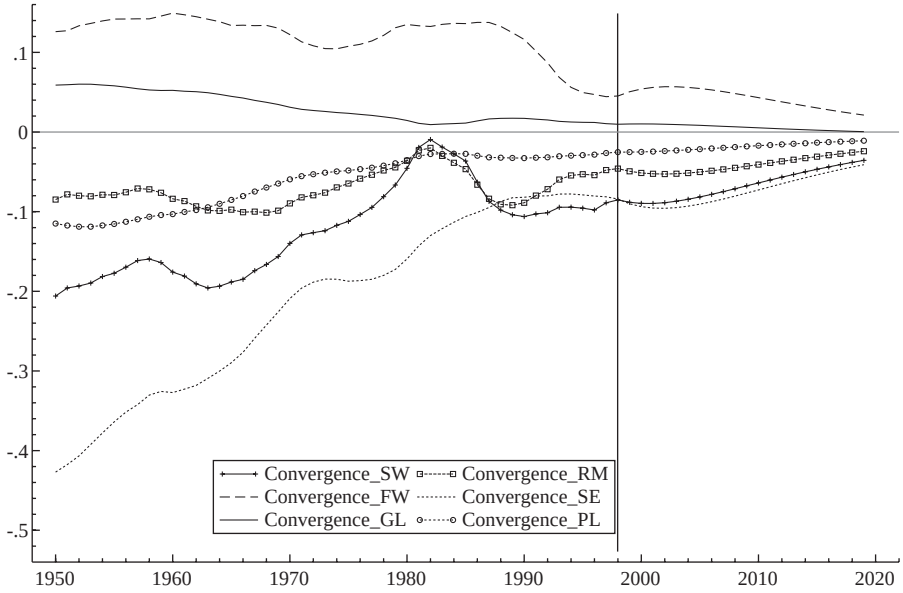

Fig. 8. Forecasts for convergence component.

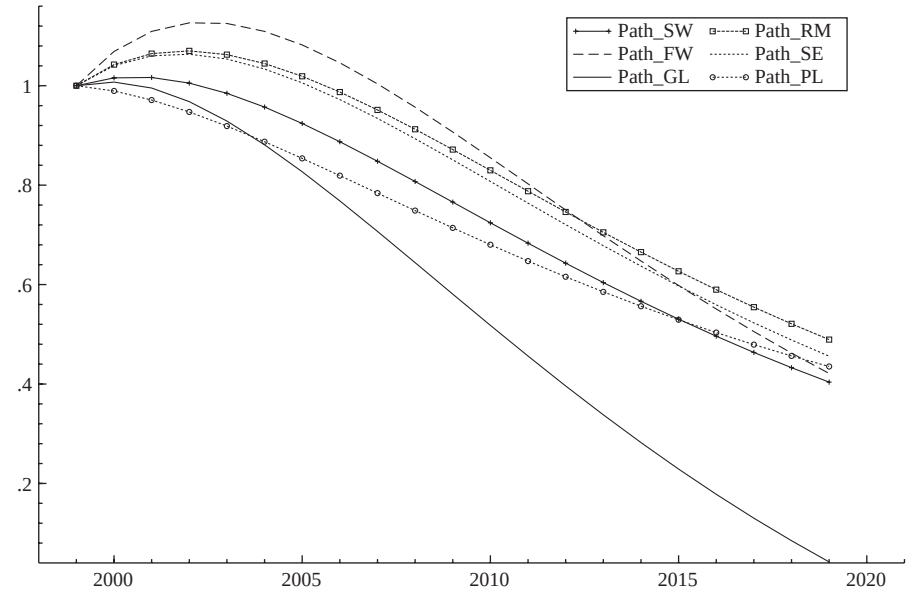

Fig. 9. Forecasts of convergence components standardised by setting to one in 1999.

also, when the convergence process stalls, allows for divergence in the short run. This is because the expected convergence path for these type of models depends not only on the convergence parameter, but also on the direction of the convergence component at the end of the sample.

Finally we contrast the behaviour of the two richest regions with that of the six poorest. To this end, we fitted a bivariate model, Eq. (9), without a convergence component, to NE and ME and extracted the trends. Fig. 10 shows the differences between each of these trends and the common trend for the six remain-

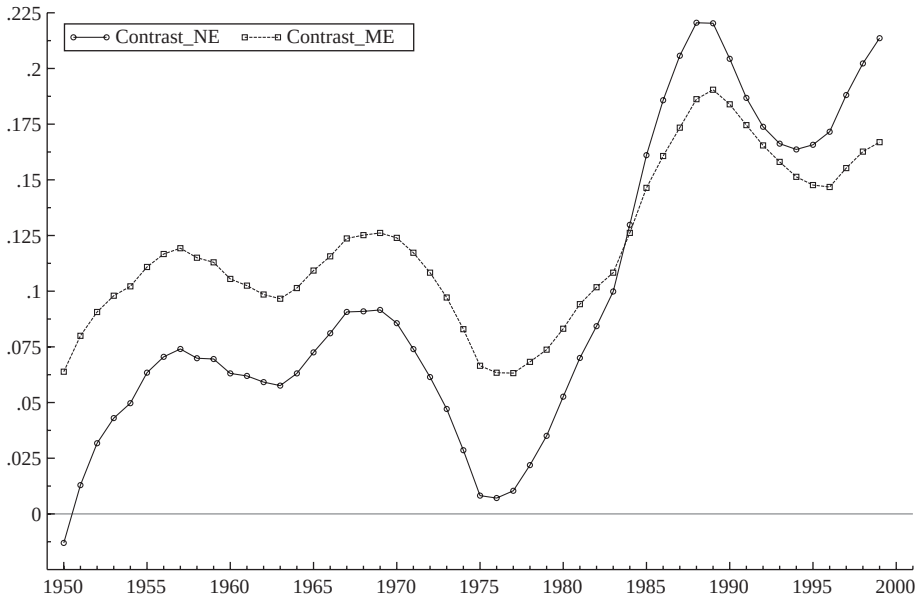

Fig. 10. Contrast of trends for NE and ME with common trend of the six poorest regions.

Table 2
ADF τ test with 4 lags and no constant

	NE	ME	GL	PL	SE	SW	RM
ME	–1.091						
GL	0.061	–0.099					
PL	–0.046	–0.421	–1.833*				
SE	–0.838	–1.511	–2.526**	–2.462**			
SW	–0.132	–0.489	–1.144	–1.114	–2.381**		
RM	–0.167	–0.687	–1.300	–1.398	–1.967**	–0.924	
FW	–1.176	–1.526	–0.748	–1.454	–2.417**	–1.069	–1.181

*means significant at 10% level (50 observations), critical value = −1.61.
**means significant at 5% level (50 observations), critical value = −1.95.

ing regions. The divergence is clear with the gap increasing markedly in the 1980s.

6. Testing convergence

The aim of convergence tests is to determine whether regions are in the process of converging. Some unit root tests are sensitive to initial conditions, rendering them of limited value for this purpose. Harvey and Bates (2003) examine a number of tests and conclude that the ADF t-test is the most satisfactory one as it is robust to initial values different from zero. Indeed, when there is no constant, its power increases the further the initial conditions are from equilibrium. Note that many researchers seem to be concerned with testing whether convergence has actually taken place. For example, Bernard and Durlauf (1996, p. 171) write that '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.' If this really were the case, then stationarity tests, rather than unit root tests, would be appropriate.

The Monte Carlo experiments in Harvey and Bates (2003) also highlight the considerable advantages of tests without the constant when absolute convergence is the hypothesis of interest. Many studies, for example Carlino and Mills (1993), carry out ADF t-tests with a time trend as well as a constant. However, the inclusion of a time trend is not only inconsistent with a convergence model but it also effectively ensures that the tests have very low power.

The usual approach to testing convergence is based on fitting models of the form[9]

$$\Delta\left(y_{i,t}-\bar{y}_{w,t}\right)=\delta_i+\pi_i\left(y_{i,t-1}-\bar{y}_{w,t-1}\right)+\sum_{r=1}^{p}\phi_{ir}\Delta\left(y_{i,t-1}-\bar{y}_{w,t-1}\right)+\eta_{it},\quad i=1,\ldots,N, \tag{35}$$

and carrying out ADF unit root tests. In the present application, $\bar{y}_{w,t}$ would usually be the US figure, rather than a simple weighted average of the individual figures in logarithms. An overall test of convergence is often carried out by combining the information in the individual ADF statistics as in Evans and Karras (1996) or Levin, Lin and Chu (2002). However, these tests do not take account of the cross-correlation between the series, thereby rendering them invalid in the present application; see the simulation evidence in O'Connell (1998). More generally, there is the issue of how useful an overall test is in the first place. As we have seen for the US, some regions may converge while others do not. Thus, individual tests may be more informative. Unfortunately, basing such tests on the equations in Eq. (35) in an attempt to determine which regions converge to the overall mean is obviously invalid since if one region does not converge but all the others do, then $y_{i,t}-\bar{y}_{w,t}$ will be nonstationary for all N and the tests tell us nothing. Our preference is therefore to use pairwise ADF tests to try to determine which regions are converging to each other.

[9] This equation does not follow from the model of Section 4 except in the homogenous case when all the π_is are the same; see Section 4.2.

Table 3
ADF τ test with 4 lags and constant

	NE	ME	GL	PL	SE	SW	RM
ME	–0.081						
GL	–0.757	–1.341					
PL	–1.681	–1.792	–1.361				
SE	–1.796	–1.785	–1.876	–1.819			
SW	–2.169	–2.356	–1.258	–2.439	–2.265		
RM	–1.689	–2.753*	–1.850	–1.653	–1.446	–1.195	
FW	–0.685	–1.299	–2.110	–0.416	–1.213	–0.954	–1.112

*means significant at 10% (50 observations), critical value = −2.60.

Tables 2 and 3 show the results of the ADF tests, with p set to four, applied to the differences between all the series. The results for the test with no constant show some support for convergence within the six poorest regions, though only 6 out of 15 of the test statistics are statistically significant at the 5% level. However, five of these involve SE. If a constant is included, only one t-statistic is statistically significant at the 10% level and this is between ME and RM! Overall, the tests appear to contain little useful information as to which regions can be grouped into convergence clusters.

7. Conclusions

Fitting a multivariate structural time series model to the eight US regions provides considerable insight into convergence, or lack of it, by focussing attention on the extraction of smooth trends. Based on these trends, a plausible working hypothesis is that all but the two richest regions, NE and ME, have displayed (absolute) convergence over the last 50 years. There are sound economic reasons why regional per capita incomes should tend to equalise. For example, in the case of the US, the recent paper by Caselli and Coleman (2001) highlights the declining role of agriculture in bringing about convergence. Explaining the underlying reasons for the recently observed divergence of the NE and ME presents a more interesting challenge.

A secondary aim of the study was to characterise the cyclical movements in US regions. This was done by fitting a similar cycle component and the result is individual regional cycles that appear entirely plausible. The first principal component of the estimated covariance matrix of the cycles accounts for the bulk of the regional movements and tracks the national cycle quite closely. Attempts to impose a single common cycle on the regional series indicated that this model is too restrictive.

The main theoretical contribution of the paper is the development of a convergence model which is then fitted to the six poorest regions. A key feature of the model is that it embodies convergence components that are able to display temporary divergence before converging to a common trend. This temporary divergence seems to be a feature of the US regions. The model not only characterises it over the period in question, but also displays it when predictions are made. These findings have important implications for the 'speed of convergence debate,' prevalent in many empirical studies of convergence: if second-order convergence components best characterise the series, no single parameter captures the speed of convergence and a simple notion of a half-life is inappropriate.

Finally, we note that unit root tests, as usually applied in the literature, are of limited value in determining which regions are converging. Building a model that provides a statistical description of the underlying movements in the economy is far more fruitful.

Acknowledgements

The work was supported by the Economic and Social Research Council (ESRC) as part of a project on Dynamic Common Factor Models for Regional Time Series, grant number L138 25 1008, based at Cambridge University. As well as being presented in Madrid, the paper was also given at the 'Common

Features in London' conference at the CASS Business School in December 2004, while an earlier version was presented at the conference on business cycles at the University of Manchester in June 2002. We are grateful to Antonio Garcia Ferrer, Esther Ruiz and the discussants for their careful reading of the paper and to all those who commented at presentations.

References

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

Bernard, A., & Durlauf, S. (1996). Interpreting tests of the convergence hypothesis. *Journal of Econometrics*, *71*, 161–173.

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

Carlino, G., & Sill, K. (2001). Regional income fluctuations: Common trends and common cycles. *Review of Economics and Statistics*, *83*, 446–456.

Carvalho, V. M., & Harvey, A. C. (2002). *Growth, cycles and convergence in US regional time series*, DAE Working paper 0221. University of Cambridge, Cambridge.

Caselli, F., & Coleman, W. J. (2001). The U.S. structural transformation and regional convergence: A re-interpretation. *Journal of Political Economy*, *109*, 584–616.

Doornik, J. A. (1999). *Ox: An object-oriented matrix language* (3rd ed.). London: Timberlake Consultants Press.

Durlauf, S., & Quah, D. (1999). The new empirics of economic growth. In J. B. Taylor, & M. Woodford (Eds.), *Handbook of Macroeconomics*, *vol. 1*. Amsterdam: Elsevier Science.

Engle, R. F., & Kozicki, S. (1993). Testing for common features. *Journal of Business and Economics Statistics*, *11*, 369–380.

Evans, P. (2000). Income dynamics in regions and countries. In G. Hess, & E. van Wincoop (Eds.), *International macroeconomics*. Cambridge: Cambridge University Press.

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

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: Cambridge University Press.

Harvey, A. C., & Bates, D. (2003). *Multivariate unit root tests and testing for convergence*, DAE Working paper 0301. Cambridge: Cambridge University Press.

Harvey, A. C., & Koopman, S. J. (1997). Multivariate structural time series models. In C. Heij, et al. (Eds.), *System dynamics in economic and financial models* (pp. 269–298). Chichester: Wiley and Sons.

Hobijn, B., & Franses, P. H. (2000). Asymptotically perfect and relative convergence of productivity. *Journal of Applied Econometrics*, *15*, 59–81.

Koopman, S. J., Shephard, N., & Doornik, J. A. (1999). Statistical algorithms for models in state space form using SsfPack 2.2 (with discussion). *Econometrics Journal*, *2*, 107–160.

Koopman, S. J., Harvey, A. C., Doornik, J. A., & Sheppard, N. (2000). *STAMP 6.0: Structural Time Series Analyser Modeller and Predictor*. London: Timberlake Consultants Ltd.

Levin, A., Lin, C. -F., & Chu, C. -S. J. (2002). Unit root tests in panel data: Asymptotic and finite sample properties. *Journal of Econometrics*, *108*, 1–24.

Loewy, M. B., & Papell, D. H. (1996). Are U.S. regional incomes converging? Some further evidence. *Journal of Monetary Economics*, *38*, 587–598.

Maddala, G. S., & Kim, I. -M. (1998). *Unit roots, co-integration, and structural change*. Cambridge: Cambridge University Press.

Nyblom, J., & Harvey, A. C. (2000). Tests of common stochastic trends. *Econometric Theory*, *16*, 176–199.

O'Connell, P. (1998). The overvaluation of purchasing power parity. *Journal of International Economics*, *44*, 1–19.

Quah, D. (1993). Galton's fallacy and the convergence hypothesis. *Scandinavian Journal of Economics*, *95*, 427–443.

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

Tomljanovich, M., & Vogelsang, T. J. (2001). Are U.S. regions converging? Using new econometric methods to examine old issues. *Empirical Economics*, *27*, 49–62.

Vahid, F., & Engle, R. F. (1993). Common trends and common cycles. *Journal of Applied Econometrics*, *8*, 341–360.

Vasco Carvalho is currently a doctoral student at the University of Chicago. The work reported in this paper was carried out while he was a research assistant at Cambridge University. His early research was on European economic history, and he is currently working on the empirics of economic growth and the application of robust control methods to macroeconomic modelling.

Andrew Harvey is Professor of Econometrics at Cambridge University and a Fellow of Corpus Christi College. He has written three books, *The Econometric Analysis of Time Series*, *Time Series Models* and *Forecasting, Structural Time Series Models and the Kalman Filter*. He has published around 100 academic papers and is a co-author of the *Structural Time Series Analyser, Modeller and Predictor* (STAMP) computer program. He is a Fellow of the Econometric Society and a Fellow of the British Academy.