University of Nottingham
UK | CHINA | MALAYSIA

research paper series

Globalisation, Productivity and Technology

Research Paper 2017/16

The Magnitude of the Task Ahead: Macro Implications of Heterogeneous Technology

By
Markus Eberhardt, Francis Teal

GEP

The Magnitude of the Task Ahead: Macro Implications of Heterogeneous Technology

Markus Eberhardt[a,b,c*] Francis Teal[c,d]

[a] *School of Economics, University of Nottingham, UK*
[b] *Centre for Economic Policy Research, UK*
[c] *Centre for the Study of African Economies, Department of Economics, University of Oxford, UK*
[d] *Institute for the Study of Labor (IZA), Germany*

October 19, 2017

Abstract: The empirical growth literature is dominated by accounting and regression methods which assume common production technology across countries. Our empirical model relaxes this assumption and further allows unobservable determinants of output (Total Factor Productivity, TFP) to differ across countries and time, while accounting for endogeneity and cross-section correlation arising from global shocks. Using manufacturing sector data for 48 economies we show that the assumption of common technology creates questionable results in accounting exercises and is rejected in our regressions. We illustrate that the erroneous choice of homogeneous technology has substantial impact on patterns and magnitudes of resulting TFP estimates.

Keywords: Cross-Country Analysis; Heterogeneous Technology; Total Factor Productivity; Common Factor Model

JEL classification: O14, O47, C23

*We are grateful to Michael Binder, Steve Bond, Francesco Caselli, Steven Durlauf, David Hendry, John Muellbauer, Hashem Pesaran, Ron Smith, Jon Temple and participants at conferences and workshops in Oxford, Stockholm, Brighton, Maastricht and Frankfurt for helpful comments and suggestions. Eberhardt acknowledges financial support from the UK Economic and Social Research Council [PTA-031-2004-00345, PTA-026-27-2048, ES/K008919/1]. Correspondence: Markus Eberhardt, School of Economics, University of Nottingham, Sir Clive Granger Building, University Park, Nottingham NG2 2RD, England; email: markus.eberhardt@nottingham.ac.uk; fax: +44 115 95 14159; phone: +44 115 846 8416.

I Introduction

"We compare this [input] index with our output index and call any discrepancy 'productivity'... It is a measure of our ignorance, of the unknown, and of the magnitude of the task that is still ahead of us." Griliches (1961, 446)

"As a careful reading of Solow (1956, 1970) makes clear, the stylized facts for which this model was developed were not interpreted as universal properties for every country in the world. In contrast, the current literature imposes very strong homogeneity assumptions on the cross-country growth process as each country is assumed to have an identical... production function."
Durlauf, Kourtellos and Minkin (2001, 929)

It is an unfortunate misconception that the canonical neoclassical growth model simultaneously developed by Solow (1956) and Swan (1956) necessarily implies that *all* economies in the world, rich or poor, industrialised or agrarian, possess the same production technology. As the above quotes show there are prominent critics of this assumption while Solow himself suggested that "whether simple parameterizations do justice to real differences in the way the economic mechanism functions in one place or another" was certainly worth 'grumbling' about (Solow, 1986, S23). Nevertheless, the notion that cross-country empirical analysis should, in case of accounting exercises, adopt or, in case of regression analysis, aim to arrive at a *common* capital coefficient of around .3 is deeply ingrained in the minds of growth economists.

Any doubters to this common technology view (c.f. common long-run equilibrium, common convergence process and common dynamics) are typically referred to a study by Gollin (2002) which provides strong evidence that the observed labour share of aggregate output of around .7 varies only little across a diverse set of countries once mismeasurement of labour income in less developed economies is accounted for. Note that Gollin (2002) does not conclude that these income shares are identical across countries, but that his data corrections result in considerable reduction in their variation and that there is no correlation between income and the remaining differences. Nevertheless, Gollin's findings are typically taken to mean that under the reasonable assumption of constant returns to scale and the perhaps somewhat less reasonable assumption of perfect competition cross-country growth and levels accounting exercises can assume a common capital coefficient of .3 and focus their energies on chipping away at other dimensions of the 'measure of our ignorance' (see Caselli, 2005; Hulten, 2010).

In this paper we revisit the issue whether technology is common across countries.[1] Using annual data for the manufacturing sector in 48 developing and developed countries for 1970 to 2002 (UNIDO, 2004) we show in panel time series regressions that technology differences are of crucial importance for understanding cross-country differences in labour productivity and their causes. Our preferred empirical models further emphasise the importance of time-series properties of output, inputs and TFP (Bond, Leblebicioglu and Schiantarelli, 2010) as well as of accounting for unobserved heterogeneity which manifests itself as cross-country correlations arising from global shocks and local spillover effects (Chudik, Pesaran and Tosetti, 2011). Like the existing cross-country growth literature our preferred empirical implementations address concerns over endogeneity and reverse causality. We find that once these empirical aspects are accounted for we obtain average technology estimates (capital coefficients) that are close to .3 with favourable residual diagnostics, whereas if we adopt the common technology assumption the estimates are substantially different from .3 and residual testing indicates serious misspecification. Our conclusion of technology heterogeneity is further supported by formal parameter homogeneity tests.

A second feature of our study is the focus on manufacturing instead of aggregate economy data. The central importance of this industrial sector for successful development has become a widely recognised 'stylised fact' in development economics. Yet in contrast to the literature on cross-country growth regressions using aggregate economy (Durlauf, Johnson and Temple, 2005) or agriculture data (Mundlak, Butzer and Larson, 2012; Eberhardt and Teal, 2013a, and references therein) there is comparatively little empirical work dedicated to the analysis of the manufacturing sector in a large cross-section of countries — with the exception of studies on the dual economy model (e.g. Martin and Mitra, 2002; Eberhardt and Teal, 2013b), and recent work by Dani Rodrik (Rodrik, 2013; McMillan, Rodrik and Verduzco-Gallo, 2014), cross-country empirical analysis at the sectoral level is typically limited to the investigation of OECD economies (Bernard and Jones, 1996a,b; Eberhardt, Helmers and Strauss, 2013). If manufacturing matters for development it is self-evidently important to learn about the pro-

[1] We refer to 'technology heterogeneity' to indicate differential production function parameters on observable inputs across countries, with unobservables captured as TFP.

duction process and its drivers in this industrial sector.

Our findings have two important implications for productivity analysis both at the sectoral and the aggregate economy level: first, like firms in different industries, different countries are characterised by different production technologies. Attempts at estimating cross-country production functions in pooled models, where by construction the same technology is imposed on all countries, are misspecified and yield biased estimates for the technology parameters and thus any TFP estimates derived from them. Second, merely allowing for technology heterogeneity is also insufficient to capture the complex production process at the country-level: in a globalising world economies interact through trade, cultural, political and other ties and at the same time are affected differentially by global phenomena such as the 1970s oil crises or the emergence of China as a major economic player. This creates a web of interdependencies within and across economies, leading to the breakdown of crucial assumptions for standard panel estimators employed in existing cross-country studies. Our empirical strategy accommodates this interplay of endogeneity, heterogeneity and commonality to provide evidence for the fundamental forces driving manufacturing development across the globe.

The remainder of the paper is structured as follows: the following Section motivates technology heterogeneity, nonstationarity and cross-section dependence, Section III lays out the empirical framework, and discusses econometric identification. Section IV introduces our data. Regression results are presented in Section V, their implication for productivity analysis is discussed in Section VI. Section VII concludes.

II Modelling technology in panel data

In this section we motivate the concerns with which we approach the estimation of cross-country production functions. We begin by motivating technology heterogeneity, then discuss salient time series and cross-section properties of the data.

The 'new growth' literature provides justification for heterogeneous technology parameters across countries. This strand of the theoretical growth literature argues that production functions differ across countries and seeks to determine the sources of this heterogeneity (Durlauf et al., 2001). This can intuitively be taken to mean that countries can choose an 'appropriate' production technology from a menu of feasible options. Representative examples from this literature include the work by Azariadis and Drazen (1990), Durlauf (1993), and Banerjee and Newman (1993). A simpler justification for heterogeneous production functions is offered by Durlauf et al. (2001), who argue that the Solow model was not intended to be valid in a common specification for *all* countries, but may still be a good way to investigate *each* country, by allowing for parameter differences *across* countries. A more formal treatment of technology heterogeneity is provided in Mundlak et al. (2012) and linked to the empirical framework we adopt here in Eberhardt and Teal (2013b).

In the long-run, macro variables such as value-added or capital stock often appear to represent 'nonstationary' processes in at least some countries (Lee, Pesaran and Smith, 1997; Pedroni, 2007). In empirical practice many studies establish that real value series typically behave as I(1) processes (Nelson and Plosser, 1982; Lee et al., 1997). Pedroni suggested that variable (non)stationarity should not be seen as a 'global' property, valid for all times, but as a "feature which describes local behaviour of the series within sample" (Pedroni, 2007, p.432).

In our general empirical model we emphasise a view of TFP as a 'measure of our ignorance' (Abramowitz, 1956), incorporating a wider set of factors that can shift the production possibility frontier (for instance "resource endowments, climate, institutions, and so on", Mankiw, Romer and Weil, 1992, p.410/1). This is in contrast to the notion of TFP as a definitive efficiency index, as commonly adopted in the microeconometric literature of productivity analysis. Furthermore, it is important to allow for the possibility that TFP is *in part* common to all countries, e.g. representing the global dissemination of non-rival scientific knowledge or global shocks, such as the 1970s oil crises. Alternatively, we can think of multiple economic, social, political and cultural ties between countries from which commonality (cross-section correlation) may arise. The individual evolution paths of the unobservables making up TFP should not be restrained to follow simple linear trends, but instead be allowed to evolve in a non-linear and even nonstationary fashion. For instance, a number of empirical papers report that their measures of TFP display nonstationarity, whether analysed at the economy level (Bond et al., 2010) or at the sectoral level (Bernard and Jones, 1996b). At the same time a highly flexible approach to empirical modelling using annual data raises the question of how business cycles influence or distort the empirical estimates (Eberhardt and Teal, 2011). All of these concerns point to the adoption of a multi-factor TFP structure that allows for common as well as country-specific elements and is uniquely suited for the analysis of productivity (Bai, 2009).

Existing empirical work has primarily concerned itself with the (potential) endogeneity of regressors in the empirical framework (e.g. Caselli, Esquivel and Lefort, 1996; Bond, Hoeffler and Temple, 2001), an issue that is given considerably more attention in the literature than the data properties or the potential misspecification of the empirical regression model. While the empirical methods adopted here can address the simultaneity between TFP shocks and input accumulation, we resort to an alternative estimation approach following Pedroni (2000) to rule out the potential of reverse causality and assure ourselves that these regressions represent production function models and not investment or labour demand equations in disguise. Thus in addition to incorporating much desirable technology heterogeneity, our empirical analysis also addresses the major concerns that have occupied the existing literature.

III Empirical Model and Identification

Our regression analysis adopts a common factor representation for a standard log-linearised Cobb-Douglas production function model. We discuss all its features in detail below. Formally, for time periods $t = 1, \ldots, T$, countries $i = 1, \ldots, N$ and inputs $m = 1, \ldots, k$ let

$$y_{it} = \boldsymbol{\beta}_i^{\boldsymbol{m}\prime} \boldsymbol{x}_{it} + u_{it} \qquad u_{it} = \alpha_i + \boldsymbol{\lambda}_i' \boldsymbol{f}_t + \varepsilon_{it} \tag{1}$$

$$x_{mit} = \pi_{mi} + \boldsymbol{\delta}_{mi}' \boldsymbol{g}_{mt} + \rho_{1mi} f_{1mt} + \ldots + \rho_{nmi} f_{nmt} + v_{mit} \tag{2}$$

$$\boldsymbol{f}_t = \boldsymbol{\varrho}' \boldsymbol{f}_{t-1} + \boldsymbol{\epsilon}_t \qquad \text{and} \qquad \boldsymbol{g}_t = \boldsymbol{\kappa}' \boldsymbol{g}_{t-1} + \boldsymbol{\epsilon}_t \tag{3}$$

where $\boldsymbol{f}_{\cdot mt}$ is a subset of $\boldsymbol{f}_t$. y_{it} represents value-added and $\boldsymbol{x}_{it}$ is a vector of observable inputs including labour and capital stock (all in logarithms). Technology parameters $\boldsymbol{\beta}_i$ can differ across countries but are assumed constant over time.[2] For unobserved TFP we employ

[2]The latter assumption is clearly restrictive, but given the focus on cross-country technology heterogeneity against the background of data restrictions in the time-series dimension we cannot relax this assumption for the *heterogeneous* regression models. For the *pooled* models we ran separate regressions using pre- and post-1985 subsamples. Estimates for POLS, CCEP and FD-OLS are virtually identical for the two sub-periods. Period estimates

a country-specific TFP level α_i in combination with a set of common factors $\boldsymbol{f}_t$ with country-specific factor loadings $\boldsymbol{\lambda}_i$. In equation (2) we provide an empirical representation of the observable inputs (here: capital, labour), which are modeled as linear functions of the unobserved common factors $\boldsymbol{f}_t$ and $\boldsymbol{g}_t$, with respective country-specific factor loadings. These factors introduces cross-section correlation in the observables and unobservables. Some of the unobserved common factors driving the variation in y_{it} in equation (1) also drive the regressors in (2). This setup induces endogeneity in that the regressors are correlated with the unobservables in the production function equation (u_{it}), making it difficult to identify $\boldsymbol{\beta}_i$ separately from $\boldsymbol{\lambda}_i$ and $\boldsymbol{\rho}_i$ (Kapetanios, Pesaran and Yamagata, 2011). Equation (3) specifies the evolution of the common factors, which includes the potential for nonstationary factors ($\boldsymbol{\varrho} = 1$, $\boldsymbol{\kappa} = 1$) and thus nonstationary inputs and output.

The most important features of this setup are (i) the potential heterogeneity in the impact of observables and unobservables on output across countries (α_i, $\boldsymbol{\beta}_i$, $\boldsymbol{\lambda}_i$), (ii) the potential nonstationarity of observables and unobservables (y_{it}, $\boldsymbol{x}_{it}$, $\boldsymbol{f}_t$, $\boldsymbol{g}_{mt}$), and (iii) the endogeneity of observable inputs created by the common factor structure. These properties have important bearings on estimation and inference in macro panel data which are at the heart of this paper. In the following we illustrate how assumptions over these aspects give rise to different empirical estimators, relying on the classification presented in Figure 1.[3]

If the data are demonstrably nonstationary, any specification choice carries implicit assumptions about the long run-equilibrium relationship in the data: any pooled regression model assumes that the cointegrating relationship is *identical across all countries* in the sample (common technology), whereas a heterogeneous model assumes the cointegrating relationship *differs across countries*. Note that if the econometrician makes the wrong decision here and estimates a pooled model for what is a heterogeneous cointegrating relationship, then the empirical results are likely spurious *by construction*.[4] Spurious results indicating serious empirical

for the FE estimator differ somewhat but 95% confidence bounds still show *considerable* overlap.

[3]We use the following abbreviations: POLS — pooled OLS; 2FE — two-way fixed effects; FD — first difference estimator; FE — country fixed effects; CCEP — Pesaran (2006) Common Correlated Effects Pooled estimator; IFE — Bai (2009) Interactive Fixed Effects estimator; CD-MG — cross-sectionally demeaned Mean Group estimator; MG — Pesaran and Smith (1995) Mean Group; GM-FMOLS — Pedroni (2000) Group-Mean Fully Modified OLS; CMG — Pesaran (2006) Common Correlated Effects Mean Group, and AMG — Augmented MG, described in detail in a Technical Appendix.

[4]This is very easy to show: since our specification choice of homogeneity — imposing a common parameter, say β — is wrong we enter linear combinations of the nonstationary observables $(\beta_i - \beta)x_{it}$ in the error terms, which are thus nonstationary by construction.

misspecification can however be detected by investigating residuals for nonstationarity or by implementing formal cointegration tests — we apply both strategies below.

Assumptions about unobservable TFP also have direct implications for specification and thus identification: if TFP is nonstationary, then we face the difficulty that the estimation of the cointegrating relationship would somehow need to account for an *unobservable* process. Again, if the econometrician makes the wrong decision here in terms of specification — common versus idiosyncratic TFP evolution or a mix of the two — then regression results may be spurious. If, on the other hand, TFP is assumed stationary, then deterministic components (year dummies, linear trends) should go a long way of accounting for its impact and we can still estimate a cointegrating relationship between observable inputs and output (Pedroni, 2007). Our empirical implementation allows us to represent different scenarios for the specification of TFP representative of our assumptions about the heterogeneity or homogeneity of TFP evolution.

One of the central focal points of the cross-country growth empirical literature over the past two decades has been the endogeneity of inputs and, closely related, potential reverse causality in the estimation equation. The former implies that the capital and labour inputs of our production function are correlated with unobservable TFP; conceptually, it seems highly plausible that technical progress does not merely affect output directly, but also affects the choice of factor inputs. Similarly for other aspects of TFP such as common shocks. Reverse causality implies that although we have written down a production function, we may run the risk of this representing a misspecified investment or labour demand equation. In the existing literature identification in the face of these difficulties is typically argued to be achieved through instrumentation, in panel models frequently employing the own-instrumentation strategy of the GMM estimators by Arellano and Bond (1991) and Blundell and Bond (1998). These estimators however assume common technology, stationary variable series, as well as cross-section independence, and their identification strategy is invalid if any of these assumptions are violated (Pesaran and Smith, 1995).[5]

Our own empirical implementation allows us to adopt a flexible approach to dealing with this endogeneity problem, in that we employ unobservable common factors $\mathbf{f}_t$ which induce the

[5]This criticism extends to the various control function estimators used in the microeconometric literature on production functions. See the discussion in Eberhardt and Teal (2011) for details.

correlation with observable inputs in all countries. Note that TFP is of course a catch-all, in that shocks such as the 1970s oil crises affect both output and inputs directly, with no means for the existing empirical analysis of production functions to distinguish this type of shock from technological progress through knowledge accumulation and diffusion. Furthermore, shocks may not always be global in nature — for instance extreme weather episodes leading to productivity shocks in only a small set of countries — so that it is important to emphasise that the common factor framework also allows us to acount for common factors which are more 'local' in their impact. In our preferred implementation the resulting endogeneity problem will be tackled by *accounting for the presence* of the unobservables in the empirical specification. In alternative implementations (i) these factors are estimated (Bai, 2009), or (ii) a time-series econometric estimation approach (fully modified OLS) corrects for the endogeneity bias arising in this setup but with more restrictive assumptions about common factors and thus TFP evolution (Phillips and Hansen, 1990; Pedroni, 2000). In order to tackle reverse causality, we will resort to a combination of the implementation dealing with the common factors and the 'fully modified OLS' approach.

Conveniently, we can employ residual diagnostic tests to investigate whether our implementation has successfully captured the systematic relationships in unobservable TFP: focusing on the time-varying aspects of TFP, there is much to be said for interdependence across countries, whereby for instance knowledge created in one country spills over imperfectly to other countries. These spillovers induce dependence between unobservable TFP across countries, and since TFP is also correlated with the observable variables of the model between labour and capital inputs across countries. By investigating whether residual series are cross-sectionally correlated we can highlight to what extent we have been able to deal with the dependence caused by the unobservable factors and thus indirectly whether we have addressed the endogeneity concern: if residuals are white noise we know that empirical results do not suffer from endogeneity bias.

As this discussion highlights, the choice between estimating a pooled and a heterogeneous model as well as the treatment of TFP in this context is not some minor specification choice but a matter of great importance. We expect to see significant differences in estimates when moving between results for pooled and heterogeneous estimators, as well as between models which make different assumptions about the nature of TFP. We expect to see that things go very wrong if we make bad specification choices: parameter estimates may have nonsensical magnitudes or turn out insignificant, residuals will be nonstationary and further diagnostic tests will indicate other serious shortcomings. This line of argument is the reason why below we also present results for estimators which we would dismiss on theoretical grounds as unreliable or biased: if the assumptions implicitly made by adopting these estimators are seriously violated, then our diagnostic tests should pick this up.

We use Figure 1 to categorise the various estimators adopted in our study and to provide some examples of previous work in the cross-country growth literature. With reference to equation (1) we also highlight the assumptions made about the TFP process in each case. The estimators assuming homogeneous technology in the upper panel of the diagram differ in their assumptions about the TFP process. The CCEP estimator by Pesaran (2006) and Bai's (2009) IFE assume that TFP evolution differs across countries but can have common elements. The former represents an augmented version of a standard fixed effects model where cross-section averages of all variables, i.e. $\bar{y}_t = N^{-1}\sum y_{it}$ and $\bar{\boldsymbol{x}}_t = N^{-1}\sum \boldsymbol{x}_{it}$, are introduced in the pooled regression to capture the unobserved common factors. In order to account for heterogeneity in the impact of these factors across countries the coefficients on the cross-section averages are allowed to differ for each country. An alternative is provided by the IFE estimator (Bai, 2009), which is in the tradition of implementations which first estimate the common factors using Principle Component Analysis and then include them in the regression equation, which is then estimated iteratively until convergence is achieved (e.g. Bai, Kao and Ng, 2009).[6] In the past one criticism of this approach focused on the necessity to employ information criteria prior to estimation to establish the number of 'relevant' common factors in the data. Recent theoretical work by Moon and Weidner (2015), however, showed that assuming too many common factors has minimal impact on the consistency of the estimator. In contrast the pooled OLS (POLS), two-way fixed effect (2FE) and first difference OLS (FD) estimators all assume common TFP evolution, captured by common year effects, but represent different assumptions about country-specific TFP levels: for 2FE and FD these are like in the CCEP assumed to differ across countries, as for instance in Islam (1995), whereas they are assumed common in the

[6] A related approach by Kneip, Sickles and Song (2012) instead combines nonparametric methods with PCA to obtain the common factors.

POLS — matching the original Mankiw et al. (1992) assumption.

Nonstationarity has different implications for this set of pooled estimators: for POLS and 2FE we assume homogeneous cointegration. Since both estimators account for time fixed effects[7] there is nothing preventing us from including unobserved TFP in this cointegrating relationship, provided it is common to all countries. If our specification choice is correct the estimates from these models under cointegration would be super-consistent, implying that endogeneity would not lead to first order bias in these models (Engle and Granger, 1987). The FD estimator is unaffected by nonstationarity, since the differencing of the estimation equation renders its observables and unobservables stationary by construction. At the same time we are prevented from making any statements about a 'long-run equilibrium' relationship from the FD estimate. The CCEP estimator *theoretically* yields consistent, but not super-consistent, estimates of β or the mean of β_i regardless of whether our choice of homogeneous cointegration is correct (Kapetanios et al., 2011). However, in practice it is often found that this estimator yields very different estimates from its Mean Group version (see below) and concerns for heterogeneity misspecification remain. The Bai (2009) IFE assumes the common factors are stationary, though since the PCA estimation is implemented on the differenced data, consistency may well extend to the nonstationary case, although to the best of our knowledge no theoretical results are available.

All models in the lower panel of the diagram allow for heterogeneous technology and are implemented in two steps: the first step represents some country-specific regression, while the second step consists of the averaging of country-specific estimates across the sample. All of these models thus represent 'Mean Group'-type estimators, named after the seminal contribution by Pesaran and Smith (1995). Again they differ in their assumptions about the TFP process, where we have to distinguish both the commonality and the nature of TFP growth over time (all models allow for different TFP levels across countries): the estimators in the first and third columns (CD-MG, AMG, CMG) allow for TFP to evolve in an unrestricted fashion, which includes the possibility of nonstationary TFP. In the latter case they can accommodate cointegration between inputs, output and TFP. These implementations however differ in their assumption about the commonality of TFP: in the CD-MG TFP evolution is assumed common to

[7] For POLS in form of year dummies, in the case of 2FE, the mathematically equivalent data transformation into deviations from the cross-section mean.

all countries in the sample, whereas in the AMG and CMG it is allowed to differ. These models are implemented by use of data in deviation from the cross-section means (CD-MG), or by augmentation of the country-specific estimation equation with cross-section averages of all variables (CMG, see Pesaran, 2006; Chudik et al., 2011; Kapetanios et al., 2011) or alternative estimated placeholders (AMG, see Bond and Eberhardt, 2013, and the Technical Appendix) — estimation is always by OLS.

The heterogeneous estimators in the second column (MG, GM-FMOLS) of the diagram in contrast assume constant TFP growth and thus stationary TFP: these estimators adopt linear trends to capture TFP evolution over time and *require* a cointegrating relationship between inputs and output. Although parameter estimates are in this case super-consistent it was found that corrections for endogeneity and dynamic misspecification — both leading to second order bias — as implemented in the 'fully modified OLS' (FMOLS) estimator are necessary in finite samples (Phillips and Hansen, 1990).

As was indicated above, for the AMG and CMG estimates we cannot rule out reverse causality, which represents a major shortcoming. In order to address this we simply adopt FMOLS versions of these estimators, thus using augmented estimation equations, where the augmentations are cross-section averages or other placeholders. This empirical strategy can address endogeneity, serial correlation and reverse causality *even in the case of nonstationary TFP*.

Inference for the pooled estimators builds on standard White heteroskedasticity-robust standard errors,[8] with the exception of the CCEP, where we employed the bootstrap. Inference in the heterogeneous parameter models follows Pesaran and Smith (1995), employing a nonparametric variance estimator to construct standard errors and t-ratios – the exception here is the Group Mean version of the FMOLS estimator, which obtains 'panel t-statistics' as $\bar{t}_{\beta^*} = N^{-1/2} \sum_i t_i$, where t_i is the t-ratio in country i and N is the number of countries.

[8]Standard errors for the capital coefficients increase to 0.05 in the POLS and to 0.11 in the FD models if we cluster by country — those for the 2FE model are unchanged since the Stata implementation we adopt clusters standard errors.

IV Data and Data Properties

For our empirical analysis we employ aggregate sectoral data for manufacturing from developed and developing countries for the period 1970 to 2002 (UNIDO, 2004) — data from the same source (albeit at a higher level of disaggregation) were recently used by Rodrik (2013) to investigate cross-country convergence in manufacturing value-added. Our sample represents an unbalanced panel of 48 countries with an average of 24 time-series observations (min: 11, max: 33).[9] Basic descriptive statistics and the sample makeup are detailed in Tables 1 and 2. The data allow us to estimate production functions with manufacturing sector value-added as output, and labour force and capital stock in manufacturing as inputs — the latter is created from data on gross fixed capital formation following the standard perpetual inventory methodology. Our focus here is on value-added specifications, though we also considered gross-output specifications, results for which can be found in Eberhardt (2009). Further discussion of the data and their construction is confined to the Technical Appendix.

In preparation for our regression analysis in Section V we carried out a range of variable unit root tests — detailed results are contained in a Technical Appendix. Despite all the problems related to panel unit root testing, as well as considering the present data dimensions and characteristics, we can conclude that these results strongly suggest that the variable series in levels are nonstationarity I(1). We further applied the Pesaran (2015) test for weak cross-section dependence to our model variables. Results in the Technical Appendix suggest that all series are subject to strong dependence.

V Regression results

Results in Table 3, Panel A, are based on estimating pooled models with variables in levels or first differences, including year dummies or in the CCEP country-specific period-averages following Pesaran (2006). Estimates for the capital coefficient in these regressions with con-

[9]We do not carry out any interpolation to fill gaps in the time series and do not account for missing observations in any way. Our preferred empirical specifications are based on heterogeneous parameter models, where arguably the unbalancedness (around 25% of observations in the balanced panel are missing) comes less to bear on the estimation results than in the homogeneous models due to the averaging of estimates. See Figure TA-1 in the Technical Appendix for details on missing observations.

stant returns to scale imposed are statistically significant at the 5% level or 1% level. For all three estimators in levels the regression diagnostics (not reported) suggest serial correlation in the error terms, while constant returns to scale are rejected at the 1% level of significance except for POLS. Further, the OLS and 2FE residuals are found to be nonstationary, suggesting the empirical results reported are potentially spurious. Cross-section dependence is present in all residual series to a greater or lesser extent, with 2FE and CCEP models rejecting weak cross-section dependence at the 5% level. The POLS results in [1] suggest that failure to account for time-invariant (TFP level) heterogeneity across countries yields biased results: at around .8 the capital coefficient is considerably inflated. Accounting for country-specific intercepts in [2] reduces these coefficient estimates somewhat. The same parameter in the CCEP results in [3] is yet lower still, around .6. The OLS regression in first differences in [4] yields quite different results: the capital coefficient is now around .3, CRS cannot be rejected, the AR(1) tests (not reported) show only first order serial correlation for this model, which is to be expected given that errors are in first differences. This echoes the favourable performance in simulation exercises to capture the *average* of a heterogeneous technology coefficient (see Bond and Eberhardt, 2013, and related online Appendix). However, recall that the first difference specification cannot be interpreted as a long-run equilibrium equation and we may well be capturing short-run (business cycle) fluctuations in these results. Nevertheless, it appears that the FD estimator obtains sound diagnostics and a theory-consistent technology estimate – this indicates that accounting for nonstationarity (of factor inputs and TFP) plays a crucial role in estimating cross-country production functions.[10]

We can make use of the year dummy coefficients derived from the pooled FD model to obtain an estimate of the common dynamic process $\hat{\mu}_t^\bullet$, an estimate of the average TFP evolution — see Technical Appendix for details. Figure 2 illustrates the evolution path of this common dynamic process for the unrestricted and CRS models. The graphs show severe slumps following the two oil shocks in the 1970s, while the 1980s and 1990s indicate considerable upward movement.[11] If we follow the 'measure of our ignorance' interpretation of TFP, then a decline

[10]Simulation exercises (Bond and Eberhardt, 2013) generally highlight the favourable performance of the FD estimator in standard nonstationary panel setups. However, while this may yield an unbiased estimate of average technology, country-specific TFP estimates are nevertheless biased if the 'true' technology differs across countries.

[11]These graphs are 'data-specific': for years where data coverage is good, this can be interpreted as 'global', whereas in later years (10 countries have data for 2001, only 2 for 2002, omitted from the graph) this interpretation collapses.

in global manufacturing TFP as evidenced in the 1970s should not be interpreted as a decline in knowledge, but a worsening global manufacturing *environment*, which seems plausible.

In the following we relax the assumption implicit in the pooled regressions that all countries possess the same production technology. At the same time, we maintain that common shocks and/or cross-sectional dependence have to be accounted for in some fashion. Unweighted averages of country parameter estimates are presented in Panel B of Table 3.[12] The t-statistics for the country-regression averages reported are measures of dispersion for the sample of country-specific estimates, following Pesaran and Smith (1995).

Our first observation regarding the averaged country results is that across all specifications the means of the capital coefficients are considerably lower than in the pooled levels models: between .2 and .5, rather than between .6 and .9.[13] Closer inspection suggests the following patterns across the heterogeneous parameter regression results: firstly, the two more restrictive specifications in [1] and [2] are misspecified. For the MG, which assumes linear TFP evolution, residual diagnostics indicate strong cross-section dependence; for the CD-MG, assuming common TFP evolution, residual appear nonstationary, so that we cannot rule out that these results are spurious. Secondly, for the AMG estimators, which account for a flexible TFP process in the estimation equation, diagnostic test results are favourable and averaged coefficients around .3. Thirdly, the results for the CMG with and without additional country trend differ considerably, with the former close to the AMG results and the latter slightly larger, around .45. Diagnostic tests however suggest that the standard CMG suffers from cross-sectionally strongly dependent residuals (see Pesaran (2015) CD test results).

Our results imply that (i) heterogeneous specifications which allow for a combination of commonality and idiosyncracy in the TFP evolution provide the closest match to the data and most favourable diagnostics; (ii) estimated capital coefficients in the preferred empirical specifications are close to .3; (iii) TFP appears to be nonstationary and thus leading to empirical

[12] Robust means weighing down outliers yield very similar results, with kernel estimates of the distribution of capital coefficients showing no influential outliers.

[13] Results presented are robust to alternative specifications (all results available on request): firstly, we estimated all models in first differences; secondly, we adopted alternative country-level deterministics (additional squared trend in the levels models, additional trend in the models in a first difference specification); thirdly, we estimated gross-output-based models including material inputs as additional covariate; and fourthly, we estimated dynamic ARDL versions of the presented static models.

misspecification in models which ignore this property;[14] (iv) our preferred results based on favourable residual diagnostics represent a close match between the Pesaran (2006) CMG and the Bond and Eberhardt (2013) AMG estimators.

These conclusions are backed up by the results for an alternative estimator, the Bai (2009) Interactive Fixed Effects, which are presented in Table 4: capital coefficients are uniformly close to .3, except in case of the single common factor model in [1], for which the estimate is close to .6. This model also displays nonstationary residuals. The diagnostics are favourable for all other models, suggesting stationary and only weakly dependent residuals, although it is notable that only the specifications with two and five common factors do not reject constant returns to scale. The relative stability of these results regardless of the number of factors included (other than the case of just a single factor) suggests that the model in column [2] assuming two factors captures the data quite well. Comparing these results with those for the AMG and CMG, it appears that the standard CMG implementation in Panel B, column [5] of Table 3, which on the basis of containing a single observed covariate (log capital stock per worker) can only capture *a single common factor* f_t, is biased upwards.

Finally, the Pedroni (2000) Group-Mean FMOLS approach for which results are presented in Table 5, provides further evidence that failure to account for nonstationary TFP leads to the collapse of the empirical estimates when analysing cross-country manufacturing production. In Panel A of the Table, where we investigate the full sample of 48 countries, we find that the standard GM-FMOLS in column [1] yields very low coefficient estimates, whereas upon inclusion of the common dynamic process in [2] and [3] or of cross-section averages in [4] and [5] we obtain results which closely match those from the previous Table of OLS-based results. Since the FMOLS methodology is robust to reverse causality this provides assurance that our AMG and CMG estimates represent production function coefficients and not misspecified investment or labour demand equations. In Panel B of Table 5 we limit the sample to 26 countries for which individual time-series unit root and stationarity (DF and KPSS) tests could not reject nonstationarity (the FMOLS approach assumes nonstationarity and cointegration),[15] to

[14]A comparison of results for the unit root analysis of the regression residuals $\hat{\varepsilon}$ and of $y - \hat{\beta}k$ or its heterogeneous technology variant (which contains $\hat{\varepsilon}$ *and* the common factors) indicates that the POLS, 2FE and CD-MG models cannot capture nonstationary TFP.

[15]We appreciate that single time series tests employed typically have low power in the present short time series, but this analysis is intended to be indicative of the remarkable robustness of our findings to a reduction in the sample to countries with *plausibly* rather than definitively I(1) variable series.

show that results do not change in any significant way.

Based on residual diagnostic our empirical results thus largely favour models with heterogeneous technology which account for a combination of heterogeneous and common TFP. The notable exception here is the (pooled) First Difference estimator, which we found relatively unaffected by the failure to explicitly model these features, likely due to the absence of integrated variables and processes once data are differenced. In our minds the fact that the FD estimator obtains a similar capital coefficient to that in the averaged AMG or CMG results is *in spite of* technology heterogeneity, and not because pooled specifications are favourable. To this end we also carried out a significant number of formal parameter homogeneity tests (see Technical Appendix) which confirmed our preference for heterogeneous technology. Since residual testing for stationarity represents a somewhat *ad hoc* cointegration test we also confirmed this property in our preferred heterogeneous model adopting the Gengenbach, Westerlund and Urbain (2016) testing procedure (for results see Technical Appendix).

Our general production function framework provides a number of insights into TFP estimation: firstly, it seems sensible to allow for maximum flexibility in the structure of the empirical TFP terms; if TFP represents a 'measure of our ignorance' then it makes sense to allow for differential TFP across countries and time, with the latter unconstrained with regard to non-stationarity.

Secondly, it further makes sense to keep an open mind about the commonality of TFP: while early empirical models (Mankiw et al., 1992; Islam, 1995) assumed common TFP growth for all countries, later studies preferred to specify differential TFP evolution across countries. We believe the arguments for commonality (non-rival nature of knowledge, spillovers, global shocks) and idiosyncracy (patents, tacit knowledge, learning-by-doing) call for an empirical specification which does not rule out either by construction.

Thirdly, following Durlauf et al. (2001) and Pedroni (2007) we argue for an empirical specification that allows for parameter heterogeneity across countries and for a shift away from the widespread focus on TFP analysis and toward an integrated treatment of the production technology *in its entirety*, including technology heterogeneity, TFP levels and growth rates.

We can illustrate the contribution of these three aspects of production technology in Figure 3, where we plot country-specific linear regressions of value-added per worker on capital stock per worker for our manufacturing data from 48 countries: in the left plot, which ignores TFP growth over time, the slopes of these production functions appear very similar, reinforcing the notion of a common production technology, whether we assume common or heterogeneous intercept terms (TFP levels) and common or heterogeneous slopes (capital coefficients). The same result obtains if we assume common TFP growth for all countries in the sample. From this we conclude that common TFP evolution in combination with either common or heterogeneous technology leads to empirical results which run counter to the macro factor share evidence, namely a capital coefficient around .7 rather than around .3.

In the right plot we adjust the value-added per worker variable for TFP evolution over time[16] and again plot the country-specific regression lines implied by a production function model. Thus allowing for heterogeneous TFP and common shocks, we can see that the fitted regression lines now provide clear evidence of technology heterogeneity, with the average capital coefficient from a heterogeneous parameter model around .3, while a pooled model still yields an inflated estimate of .79. From this we can conclude that heterogeneous TFP evolution alone yields results in conflict with the macro data, whereas the combination of heterogeneous technology and heterogeneous TFP evolution yields a global average of .3.

VI TFP in a heterogeneous technology world

What are the implications of homogeneity misspecification for estimated TFP levels and growth rates? In the following we provide some insights into the resulting patterns of TFP growth and introduce a new approach to estimate TFP levels which is necessitated by the adoption of a heterogeneous technology model. In both cases we try to establish whether the choice between homogeneous and heterogeneous technology makes a substantial difference to TFP measurement.

In the top left plot of Figure 4 we compare the distribution of the annual TFP growth estimates from growth accounting (dashed transparent histogram) and our preferred panel time series

[16] We compute $\text{ly}^{\text{adj}} = \text{ly} - \hat{c}_i t - \hat{d}_i \hat{\mu}^{\bullet}_t$ following equation (4) below, using the results for the empirical model in Table 3, Panel B, column [3].

regression (grey histogram). While both distributions look Gaussian, it is obvious that the accounted TFP growth rates are substantially greater in range. The top right plot in the same Figure fits a linear regression line (with 90% confidence bands) for the annual TFP growth rates against value-added per worker (in logs). While the *estimated* TFP growth rates from the preferred heterogeneous estimator seem to display a negative relationship with output, a tendency which disappears if we omit the top and bottom 5% of the distribution in the bottom right plot or if we employ total period averages of TFP growth and value-added per worker in the bottom left plot, the *accounted* TFP growth rates consistently display a positive relationship regardless of censoring or averaging.

We can draw two conclusions from this analysis: firstly, the range and variance of the common technology TFP growth estimates are artificially inflated, thus providing increased likelihood of statistically significant results in further 'TFP regressions.' Secondly, under the assumption of common technology these TFP growth series are clearly linked to the *level* of development, with richer countries enjoying higher TFP growth.

A further implication of a shift from common to heterogeneous technology is that we require a new methodology to arrive at TFP level estimates from our preferred country-level regression models: from these regressions we can obtain estimates for the intercept, technology parameters, idiosyncratic and common trend coefficients or the parameters on the cross-section averages for AMG and CMG specifications, respectively. One may be tempted to view the coefficients on the intercepts as TFP level estimates, just like in the pooled fixed effects case. However, once we allow for heterogeneity in the slope coefficients, the interpretation of the intercept as an estimate for base-year TFP level is no longer valid, as was already recognised by Bernard and Jones (1996a). In order to illustrate our case, we employ a simple linear relationship between value-added and capital where the contribution of TFP *growth* has already been accounted for (see Equation (4) below).

In Figure 5 we provide scatter plots for 'adjusted' log value-added per worker (y-axis) against log capital per worker (x-axis) as well as a fitted regression line for these observations in each of the following four countries: in the upper panel France (circles) and Belgium (triangles), in the lower panel South Korea (circles) and Malaysia (triangles). The 'adjustment' is based on the country-specific estimates from the AMG regression in Table 3, Panel B: we compute

$$y_{it}^{adj} = y_{it} - \hat{c}_i t - \hat{d}_i \hat{\mu}_t^{\bullet} \tag{4}$$

where $\hat{c}_i$ and $\hat{d}_i$ are the country-specific estimates for the linear trend term and the common dynamic process respectively. We then plot this variable against log capital per worker for each country separately. This provides a visual equivalent of the estimates for the capital coefficient (slope) and a candidate TFP level estimate (intercept) in the country regression.

The upper panel of Figure 5 shows two countries (France, Belgium) with virtually identical capital coefficient estimates (slopes). The in-sample fitted regression line is plotted as a solid line, the out-of-sample extrapolation toward the y-axis is plotted in dashes. The country-estimates for the intercepts can be interpreted as TFP levels, since these countries have very similar capital coefficient estimates ($\hat{b}_{FRA} \approx \hat{b}_{BEL} \approx \hat{b}$). In this case, the graph represents the linear model $y_{it}^{adj} = \hat{a}_i + \hat{b}\log(K/L)_{it}$, where $\hat{a}_i$ possesses the *ceteris paribus* property. In contrast, the lower panel shows two countries (Malaysia, South Korea) which exhibit very different capital coefficient estimates. In this case $\hat{a}_i$ cannot be interpreted as possessing the *ceteris paribus* quality since $\hat{b}_{MYS} \neq \hat{b}_{KOR}$: *ceteris non paribus*, or as Bernard and Jones (1996a) put it: 'comparing apples to oranges.' In the graph we can see that Malaysia has a considerably higher intercept term than South Korea, even though the latter's observations lie above those of the former at any given point in time. This illustrates that once technology parameters in the production function differ across countries the regression intercept can no longer be interpreted as a TFP-level estimate.

We can suggest an alternative measure for TFP-levels which is robust to parameter heterogeneity. Referring back to the scatter plots in Figure 5, we marked the base-year level of log capital per worker by vertical lines for each of the four countries. We suggest to use the locus where the solid (in-sample) regression line hits the vertical base-year capital stock level as an indicator of TFP-level in the base year. These *adjusted* base-year and final-year TFP-levels are thus

$$\hat{a}_i + \hat{b}_i \log(\mathrm{K/L})_{0,i} \qquad \text{and} \qquad \hat{a}_i + \hat{b}_i \log(\mathrm{K/L})_{0,i} + \hat{c}_i \tau + \hat{d}_i \hat{\mu}_{\tau}^{\bullet} \tag{5}$$

respectively, where $\log(K/L)_{0,i}$ is the *country-specific* base-year value for capital per worker (in logs), τ is the total period for which country i is in the sample and $\hat{\mu}^{\bullet}_{\tau}$ is the accumulated common TFP growth for this period τ with the country-specific parameter $\hat{d}_i$ — it is easy to see that the intercept-problem only has bearings on TFP-*level* estimates.

Table 6 provides details on absolute rank differences implied by TFP level rankings for accounting ('Levels') and regression ('2FE', 'AMG' and 'CMG') exercises. These descriptives indicate the very substantial differences arising from TFP levels obtained from common versus heterogeneous technology models.

VII Concluding remarks

In this paper we investigated how manufacturing sector technology differences across countries can be modelled empirically. We adopted an encompassing framework which allows for the possibility that the impact of observable and unobservable inputs on output differs across countries, as well as for nonstationary evolution of these processes. Our regression framework enabled us to model a number of characteristics which are likely to be prevalent in manufacturing data from a diverse set of countries: firstly, we allowed for technology heterogeneity across countries. Empirical results are confirmed by formal testing procedures to suggest that technology parameters in manufacturing production indeed differ across countries. This finding supports earlier work using aggregate economy data (Durlauf, 2001; Pedroni, 2007): if production technology differs in cross-country manufacturing, aggregate economy technology is unlikely to be homogeneous.

Secondly, we allowed for unobserved common factors to drive output, but with differential impact across countries, thus inducing cross-section dependence. These common factors are visualised by our common dynamic process, which follows patterns over the 1970-2002 sample period that match historical events. The interpretation of this common dynamic process $\hat{\mu}^{\bullet}_{t}$ would be that for the manufacturing sector *similar factors* drive production in all countries, *albeit to a different extent*. This is equivalent to suggesting the 'global tide' of innovation can 'lift all boats', but that technology transfer from developed to developing countries is dependent on the recipient's production technology and absorptive capacity, among other things.

Thirdly, our empirical setup allows for a type of endogeneity whereby unobservables driving output are also driving the evolution of inputs. This leads to an identification problem, in that standard panel estimators cannot identify the parameters on the observable inputs as distinct from the impact of unobservables. Monte Carlo simulations (Bond and Eberhardt, 2013) have highlighted the ability of the CMG and AMG estimates to deal with this problem successfully and our empirical results indicate parity between these two heterogeneous panel estimators. Furthermore, additional analysis confirms that the empirical results are robust to the use of an alternative panel time-series econometric approach which further addresses reverse causality. Standard practices to deal with endogeneity (Arellano and Bond, 1991; Blundell and Bond, 1998) are only appropriate in a stationary framework with homogeneous technology (Pesaran and Smith, 1995). Adopting a nonstationary panel econometric approach that accounts for cross-section dependence in our view is a sound empirical strategy to address both these concerns and should be applied more widely to cross-country productivity-analysis.

Our analysis represents a step toward making cross-country empirics relevant to individual countries by moving away from empirical results that characterise the *average country* and toward a deeper understanding of the *differences across countries*, a notion which is clearly echoed elsewhere in the literature (Temple, 1999; Durlauf, 2001; Durlauf et al., 2001, 2005). Cross-country regressions of time averages, in the empirical tradition of Barro (1991) and Mankiw et al. (1992), emphasise the variation in the data across countries ('between variation') and implicitly assume that the processes driving capital accumulation in, say, the United States and Malawi are the same, and that at a distant point in time the latter can feasibly reach the capital-labour ratio of the former to achieve the same level of development. However, development is an evolution over time which requires that apart from recognising the potential for differences across countries we analyse the individual evolution paths of countries over time (emphasising the 'within variation' in the data). The empirical methods used in this paper enable us to incorporate all of these concerns within one unifying empirical framework. A second important conclusion from this study is that the key to understanding cross-country differences in income is not exclusively linked to understanding TFP differences, but requires a careful concern for differences in production technology. Since modelling production technology as heterogeneous across countries requires an entirely different set of empirical methods we have focused on developing this aspect in the present paper and have left empirical testing of rival hypotheses about the patterns and sources of technological differences for future research.

References

Abramowitz, Moses (1956). "Resource and output trends in the United States since 1870." *American Economic Review*, Vol. 46(2): 5–23.

Arellano, Manuel and Bond, Stephen R. (1991). "Some tests of specification for panel data." *Review of Economic Studies*, Vol. 58(2): 277–297.

Azariadis, Costas and Drazen, Allan (1990). "Threshold Externalities in Economic Development." *Quarterly Journal of Economics*, Vol. 105(2): 501–26.

Bai, Jushan (2009). "Panel Data Models with Interactive Fixed Effects." *Econometrica*, Vol. 77(4): 1229–1279.

Bai, Jushan, Kao, Chihwa and Ng, Serena (2009). "Panel cointegration with global stochastic trends." *Journal of Econometrics*, Vol. 149(1): 82–99.

Banerjee, Abhijit and Newman, Andrew (1993). "Occupational Choice and the Process of Development." *Journal of Political Economy*, Vol. 101.

Barro, Robert J. (1991). "Economic growth in a cross-section of countries." *Quarterly Journal of Economics*, Vol. 106(2): 407–443.

Bernard, Andrew B. and Jones, Charles I. (1996a). "Comparing Apples to Oranges: Productivity Convergence and Measurement across Industries and Countries." *American Economic Review*, Vol. 86(5): 1216–38.

Bernard, Andrew B. and Jones, Charles I. (1996b). "Productivity across Industries and Countries: Time Series Theory and Evidence." *Review of Economics and Statistics*, Vol. 78(1): 135–46.

Blundell, Richard and Bond, Stephen R. (1998). "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, Vol. 87(1): 115–143.

Bond, Stephen R. and Eberhardt, Markus (2013). "Accounting for unobserved heterogeneity in panel time series models." Unpublished mimeo, November.

Bond, Stephen R., Hoeffler, Anke and Temple, Jonathan (2001). "GMM Estimation of Empirical Growth Models." CEPR Discussion Paper #3048.

Bond, Stephen R., Leblebicioglu, Asli and Schiantarelli, Fabio (2010). "Capital Accumulation and Growth: A New Look at the Empirical Evidence." *Journal of Applied Econometrics*, Vol. 15(7): 1073–1099.

Caselli, Francesco (2005). "Accounting for Cross-Country Income Differences." In: Philippe Aghion and Steven Durlauf (Editors), "Handbook of Economic Growth," Vol. 1 of *Handbook of Economic Growth*, chapter 9 (Elsevier), pp. 679–741.

Caselli, Francesco, Esquivel, Gerardo and Lefort, Fernando (1996). "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics." *Journal of Economic Growth*, Vol. 1(3): 363–89.

Chudik, Alexander, Pesaran, M. Hashem and Tosetti, Elisa (2011). "Weak and Strong Cross Section Dependence and Estimation of Large Panels." *Econometrics Journal*, Vol. 14(1): C45–C90.

Costantini, Mauro and Destefanis, Sergio (2009). "Cointegration analysis for cross-sectionally dependent panels: The case of regional production functions." *Economic Modelling*, Vol. 26(2): 320–327.

Durlauf, Steven N. (1993). "Nonergodic Economic Growth." *Review of Economic Studies*, Vol. 60(2): 349–66.

Durlauf, Steven N. (2001). "Manifesto for a growth econometrics." *Journal of Econometrics*, Vol. 100(1): 65–69.

Durlauf, Steven N., Johnson, Paul A. and Temple, Jonathan R.W. (2005). "Growth Econometrics." In: Philippe Aghion and Steven Durlauf (Editors), "Handbook of Economic Growth," Vol. 1 of *Handbook of Economic Growth*, chapter 8 (Elsevier), pp. 555–677.

Durlauf, Steven N., Kourtellos, Andros and Minkin, Artur (2001). "The local Solow growth model." *European Economic Review*, Vol. 45(4-6): 928–940.

Eberhardt, Markus (2009). "Modelling Technology in Agriculture and Manufacturing using Cross-Country Panel Data." DPhil thesis, St John's College, University of Oxford, available online at https://ora.ox.ac.uk/objects/uuid:d60f62f5-43e2-4473-b899-f4358d758e1e.

Eberhardt, Markus, Helmers, Christian and Strauss, Hubert (2013). "Do spillovers matter when estimating private returns to R&D?" *The Review of Economics and Statistics*, Vol. 95(2): 436–448.

Eberhardt, Markus and Teal, Francis (2011). "Econometrics for Grumblers: A New Look at the Literature on Cross-Country Growth Empirics." *Journal of Economic Surveys*, Vol. 25(1): 109–155.

Eberhardt, Markus and Teal, Francis (2013a). "No Mangoes in the Tundra: Spatial Heterogeneity in Agricultural Productivity Analysis." *Oxford Bulletin of Economics and Statistics*, Vol. 75: 914–939.

Eberhardt, Markus and Teal, Francis (2013b). "Structural Change and Cross-Country Growth Empirics." *World Bank Economic Review*, Vol. 27: 229–71.

Engle, R.F. and Granger, C.W.J. (1987). "Cointegration and Error correction: representations, estimation and testing." *Econometrica*, Vol. 55(2): 252–276.

Gengenbach, Christian, Westerlund, Joakim and Urbain, Jean-Pierre (2016). "Error Correction Testing in Panels with Common Stochastic Trends." *Journal of Applied Econometrics*, Vol. 31: 982–1004.

Gollin, Douglas (2002). "Getting Income Shares Right." *Journal of Political Economy*, Vol. 110(2): 458–474.

Griliches, Zvi (1961). "Comment on *An Appraisal of Long-Term Capital Estimates: Some Reference Notes* by Daniel Creamer." Output, Input, and Productivity Measurement (NBER), pp. 446–9.

Hulten, Charles R. (2010). "Growth Accounting." Vol. 2 of *Handbook of the Economics of Innovation*, chapter 23. pp. 987–1031.

Islam, Nazrul (1995). "Growth Empirics: A Panel Data Approach." *Quarterly Journal of Economics*, Vol. 110(4): 1127–70.

Kapetanios, George, Pesaran, M. Hashem and Yamagata, Takashi (2011). "Panels with Nonstationary Multifactor Error Structures." *Journal of Econometrics*, Vol. 160(2): 326–348.

Kneip, Alois, Sickles, Robin C and Song, Wonho (2012). "A new panel data treatment for heterogeneity in time trends." *Econometric Theory*, Vol. 28(3): 590–628.

Lee, Kevin, Pesaran, M. Hashem and Smith, Ron P. (1997). "Growth and Convergence in a Multi-country Empirical Stochastic Solow Model." *Journal of Applied Econometrics*, Vol. 12(4): 357–92.

Mankiw, N. Gregory, Romer, David and Weil, David N. (1992). "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics*, Vol. 107(2): 407–437.

Martin, Will and Mitra, Devashish (2002). "Productivity Growth and Convergence in Agriculture versus Manufacturing." *Economic Development and Cultural Change*, Vol. 49(2): 403–422.

McMillan, Margaret, Rodrik, Dani and Verduzco-Gallo, Inigo (2014). "Globalization, Structural Change and Productivity Growth, with an update on Africa." *World Development*, Vol. 63: 11–32.

Moon, Hyungsik Roger and Weidner, Martin (2015). "Linear regression for panel with unknown number of factors as interactive fixed effects." *Econometrica*, Vol. 83(4): 1543–1579.

Mundlak, Yair, Butzer, Rita and Larson, Donald F. (2012). "Heterogeneous technology and panel data: The case of the agricultural production function." *Journal of Development Economics*, Vol. 99(1): 139–149.

Nelson, Charles R. and Plosser, Charles R. (1982). "Trends and random walks in macroeconomic time series: some evidence and implications." *Journal of Monetary Economics*, Vol. 10(2): 139–162.

Pedroni, Peter (2000). "Fully modified OLS for heterogeneous cointegrated panels." In: Badi H. Baltagi (Editor), "Nonstationary panels, cointegration in panels and dynamic panels," (Amsterdam: Elsevier).

Pedroni, Peter (2007). "Social capital, barriers to production and capital shares: implications for the importance of parameter heterogeneity from a nonstationary panel approach." *Journal of Applied Econometrics*, Vol. 22(2): 429–451.

Pesaran, M. Hashem (2006). "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica*, Vol. 74(4): 967–1012.

Pesaran, M. Hashem (2007). "A simple panel unit root test in the presence of cross-section dependence." *Journal of Applied Econometrics*, Vol. 22(2): 265–312.

Pesaran, M. Hashem (2015). "Testing Weak Cross-Sectional Dependence in Large Panels." *Econometric Reviews*, Vol. 34(6-10): 1089–1117.

Pesaran, M. Hashem and Smith, Ron P. (1995). "Estimating long-run relationships from dynamic heterogeneous panels." *Journal of Econometrics*, Vol. 68(1): 79–113.

Phillips, Peter C. B. and Hansen, Bruce E (1990). "Statistical Inference in Instrumental Variables Regression with I(1) Processes." *Review of Economic Studies*, Vol. 57(1): 99–125.

Rodrik, Dani (2013). "Unconditional Convergence in Manufacturing." *Quarterly Journal of Economics*, Vol. 128(1): 165.

Solow, Robert. M. (1956). "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics*, Vol. 70(1): 65–94.

Solow, Robert M (1986). "Unemployment: Getting the Questions Right." *Economica*, Vol. 53(210): S23–34.

Swan, Trevor W. (1956). "Economic Growth and Capital Accumulation." *Economic Record*, Vol. 32(2): 334–61.

Temple, Jonathan (1999). "The New Growth Evidence." *Journal of Economic Literature*, Vol. 37(1): 112–156.

UNIDO (2004). "UNIDO Industrial Statistics 2004." Online database, Vienna: UNIDO, united nations Industrial development organisation.

Tables and Figures

Table 1: Descriptive statistics

VARIABLES IN LEVEL TERMS

Variable	obs	mean	median	std. dev.	min.	max.
levels						
value-added	1,194	5.47E+10	9.04E+09	1.78E+11	1.76E+07	1.50E+12
labour	1,194	1,469,186	502,214	2,924,524	5,552	1.97E+07
capital	1,194	1.32E+11	2.61E+10	3.12E+11	5.78E+07	2.27E+12
logs						
value-added	1,194	22.70	22.93	2.15	16.68	28.04
labour	1,194	12.92	13.13	1.79	8.62	16.79
capital	1,194	23.72	23.98	2.22	17.87	28.45
annual growth rate						
value-added	1,128	3.9%	3.5%	12.3%	-78.3%	92.7%
labour	1,128	1.7%	0.7%	8.1%	-38.8%	78.1%
capital	1,128	4.1%	3.1%	4.4%	-2.4%	47.8%

VARIABLES IN PER WORKER TERMS

Variable	obs	mean	median	std. dev.	min.	max.
levels						
value-added	1,194	76,932	45,865	72,843	2,007	346,064
capital	1,194	25,305	17,867	19,385	1,660	91,011
logs						
value-added	1,194	9.78	9.79	0.92	7.41	11.42
capital	1,194	10.80	10.73	1.00	7.60	12.75
annual growth rate						
value-added	1,128	2.2%	2.5%	10.8%	-90.3%	74.4%
capital	1,128	2.5%	2.5%	7.9%	-68.0%	45.4%

Notes: We report the descriptive statistics for value-added, labour and capital stock for $N = 48$ countries and $n = 1,194$ ($n = 1,128$) observations in the levels (growth) specification. Monetary values are in real US$ (base year 1990). Labour is in number of workers.

Table 2: Sample of countries and number of observations

Country	Code	*levels*	*FD*	$t = 1$	$t = T$	Gaps
Australia	AUS	20	17	1970	1993	2
Austria	AUT	30	28	1970	2000	1
Belgium	BEL	28	27	1970	1997	-
Bangladesh$^\sharp$	BGD	14	12	1970	1992	1
Bolivia$^\natural$	BOL	11	10	1987	1997	-
Barbados	BRB	26	25	1970	1995	-
Canada	CAN	21	20	1970	1990	
Chile	CHL	25	24	1974	1998	-
Colombia	COL	30	29	1970	1999	-
Cyprus	CYP	33	32	1970	2002	-
Ecuador	ECU	30	29	1970	1999	-
Egypt	EGY	26	25	1970	1995	-
Spain	ESP	26	25	1970	1995	-
Finland	FIN	28	26	1970	2000	1
Fiji	FJI	25	24	1970	1994	-
France	FRA	26	25	1970	1995	-
United Kingdom	GBR	23	22	1970	1992	-
Guatemala$^\sharp$	GTM	16	15	1973	1988	-
Hungary	HUN	26	25	1970	1995	-
Indonesia	IDN	26	25	1970	1995	-
India	IND	32	31	1970	2001	-
Ireland	IRL	22	21	1970	1991	-
Iran	IRN	24	22	1970	2001	-
Israel$^\natural$	ISR	13	12	1989	2001	-
Italy	ITA	31	30	1970	2000	-
Korea	KOR	32	31	1970	2001	-
Sri Lanka	LKA	20	17	1970	2000	2
Luxembourg	LUX	23	22	1970	1992	-
Morocco$^\sharp$	MAR	17	16	1985	2001	-
Mexico$^\sharp$	MEX	16	14	1984	2000	1
Malta	MLT	32	31	1970	2001	-
Malaysia	MYS	28	25	1970	2001	2
Netherlands	NLD	24	23	1970	1993	-
Norway	NOR	32	31	1970	2001	-
New Zealand	NZL	21	20	1970	1990	-
Panama	PAN	30	28	1970	2000	1
Philippines	PHL	26	25	1970	1995	-
Poland	POL	31	30	1970	2000	-
Portugal	PRT	31	30	1970	2000	-
Senegal$^\sharp$	SEN	17	14	1970	1990	2
Singapore	SGP	33	32	1970	2002	-
Sweden$^\sharp$	SWE	18	17	1970	1987	-
Swaziland	SWZ	24	22	1970	1995	1
Tunisia	TUN	21	19	1970	1997	1
Turkey	TUR	27	25	1970	1997	1
United States	USA	26	25	1970	1995	-
Venezuela	VEN	26	24	1970	1998	1
Zimbabwe	ZWE	27	26	1970	1996	-
Countries		48	48			
Observations		1,194	1,128			

Notes: Countries highlighted in bold represent the sample used in the second set of GM-FMOLS regressions, Table 5 Panel B ($n = 644$, $N = 26$). 'levels' and 'FD' refer to specifications in levels and first differences, respectively; $t = 1$ and $t = T$ report the start and end years of the country series; 'gaps' indicates the number of gaps in the data series.
$\natural$ These countries had to be omitted to compute the Pesaran (2015) CD test for regression models in levels, due to the lack of overlap between the omitted series and the remainder of the sample. $\sharp$ These countries had to be omitted in addition to those already identified to compute the CD test for regression models in first differences.

Table 3: Main Regression Results

PANEL A: POOLED MODELS

	[1]	[2]	[3]	[4]
estimator	**POLS**	**2FE**	**CCEP**	**FD**
dependent variable	ly	ly	ly	Δly
log capital pw	0.7895 [0.011]**	0.6752 [0.066]**	0.5823 [0.037]**	
Δlog capital pw				0.3195 [0.089]**
Diagnostics				
CRS: p-value	.96	.00	.00	.72
$(y-\hat{\beta}k)$ I(1): p-value	.99	.99	.99	.00
$\hat{\varepsilon}$ I(1): p-value	1.00	.78	.00	.00
$\hat{\varepsilon}$ CD: p-value	.15	.05	.03	.39
RMSE	.462	.135	.113	.103

PANEL B: HETEROGENEOUS MODELS (AVERAGE ESTIMATES)

	[1]	[2]	[3]	[4]	[5]	[6]
estimator	**MG**	**CD-MG**	**AMG**	**AMG**	**CMG**	**CMG**
dependent variable	ly	ly$^\flat$	ly-$\hat{\mu}^\bullet_t$	ly	ly	ly
log capital pw	0.1789 [0.081]*	0.5295 [0.056]**	0.2896 [0.074]**	0.2982 [0.081]**	0.4663 [0.070]**	0.3125 [0.085]**
common dynamic process				0.8787 [0.202]**		
country trend	0.0174 [0.003]**		0.0001 [0.003]	0.0023 [0.004]		0.0108 [0.004]**
Diagnostics						
CRS: p-value	.90	.20	.99	.96	.05	.98
$(y-\hat{\beta}_i k)$ I(1): p-value	.99	.66	.99	.99	.51	.99
$\hat{\varepsilon}$ I(1): p-value	.00	.60	.00	.00	.00	.00
$\hat{\varepsilon}$ CD: p-value	.00	.25	.96	.30	.02	.82
RMSE	.100	.123	.097	.091	.100	.088

Notes: Regressions are for N=48 countries, n=1,194 (n=1,128) observations in the levels (first difference) specifications. Values in brackets are White heteroskedasticity-consistent standard errors in Panel A, except for [3] where we present bootstrapped (100 replication) standard errors; and standard errors following Pesaran and Smith (1995) in Panel B. We indicate statistical significance at the 5% and 1% level by * and ** respectively. Intercept estimates as well as average estimates on cross-section averages in Model [3] of Panel A and Models [5] and [6] of Panel B are omitted to save space.

Dependent variable: ly — log value-added per worker. ly$^\flat$ — log value added per worker in deviation from the cross-section mean (dto. for capital stock pw). Δly — growth rate of value-added (per worker). $\hat{\mu}^\bullet_t$ in Panel B is derived from the year dummy coefficients of a pooled regression (CRS imposed) in first differences (FD) as described in the main text. Models [1], [2] and [4] in Panel A contain $T-1$ year dummies (for [4] in first differences).

For all diagnostic tests (except RMSE) we report p-values: (i) The null hypothesis for the 'CRS' Wald tests is constant returns. (ii) '$(y-\hat{\beta}_i k)$ I(1)' reports analysis of regression residuals *incorporating TFP*, using a Pesaran (2007) CIPS test with 2 lags, null of nonstationarity (full results available on request). For this and the following CIPS test we adopted a pragmatic approach in setting the lags equal to 2: shorter lags may be insufficient to capture serial correlation, longer lags will demand too much from the data given the very moderate time series dimension (and further force us to drop country series). (iii) '$\hat{\varepsilon}$ I(1)' reports results for a Pesaran (2007) CIPS test with 2 lags, null of nonstationarity (full results available on request). (iv) The Pesaran (2015) CD test has the null of cross-sectional weak dependence. Due to data restrictions (unbalanced panel with missing observations) we are forced to drop 2 (8) countries from the sample to compute this test for the levels (FD) residuals. (v) RMSE is the root mean squared error.

Table 4: Homogeneous Models using Bai's (2009) IFE

	[1]	[2]	[3]	[4]	[5]
estimator	**IFE**	**IFE**	**IFE**	**IFE**	**IFE**
dependent variable	ly	ly	ly	ly	ly
number of factors	1	2	3	4	5
log capital pw	0.5645	0.3380	0.3251	0.3378	0.2654
	[.0983]***	[.0770]***	[.0637]***	[.0847]***	[.0557]***
Diagnostics					
CRS: p-value	0.10	0.47	0.00	0.00	0.29
$(y-\hat{\beta}_i k)$ I(1): p-value	0.95	0.33	0.38	0.33	0.63
$\hat{\varepsilon}$ I(1): p-value	0.37	0.01	0.00	0.00	0.00
$\hat{\varepsilon}$ CD: p-value	0.23	0.13	0.18	0.22	0.31
RMSE	0.112	0.096	0.086	0.077	0.065

Notes: The results presented are for the Bai (2009) Interactive Fixed Effects (IFE) estimator. All regression models absorb country and time fixed effects – results without these fixed effects are broadly similar (available on request). Values in brackets are absolute standard errors clustered at the country-level. Results using a bootstrap procedure are broadly in line with those presented above albeit less precise – this is not surprising given the unbalanced nature of the panel. All models were estimated in `Stata` using the `regife` written by Matthieu Gomez. See notes to Table 3.

Table 5: Heterogeneous Models using FMOLS (average estimates)

PANEL A: FULL SAMPLE (N=48)

	[1]	[2]	[3]	[4]	[5]
estimator: **‘ ’-FMOLS**	**GM**	**AMG**	**AMG**	**CMG**	**CMG**
dependent variable	ly	ly-$\hat{\mu}_t^\bullet$	ly	ly	ly
log capital pw	0.1663	0.2659	0.2937	0.5544	0.3042
	[0.084]	[0.080]**	[0.092]**	[0.069]**	[0.091]**
common process			0.8977		
			[0.257]**		
country trends	0.0171	0.0004	0.0014		0.0108
	[0.003]**	[0.003]	[0.005]		[0.004]**
Panel-t statistics, diagnostics					
capital pw	18.29	14.73	15.36	40.59	15.88
trends	24.94	18.93	12.71		20.70
RMSE	.099	.096	.090	.103	.088

PANEL B: I(1) SAMPLE (N=26)

	[1]	[2]	[3]	[4]	[5]
estimator: **‘ ’-FMOLS**	**GM**	**AMG**	**AMG**	**CMG**	**CMG**
dependent variable	ly	ly-$\hat{\mu}_t^\bullet$	ly	ly	ly
log capital pw	0.0816	0.2675	0.2784	0.5528	0.2485
	[0.064]	[0.065]**	[0.090]**	[0.075]**	[0.079]**
common process			0.8034		
			[0.174]**		
country trends	0.0179	-0.0012	0.0019		0.0108
	[0.003]**	[0.003]	[0.005]		[0.004]*
Panel-t statistics, diagnostics					
capital pw	11.45	10.37	9.97	34.96	10.16
trends	23.28	14.63	10.56		17.10
RMSE	.071	.068	.065	.080	.062

Notes: The results in [1] are for the Pedroni (2000) Group-Mean FMOLS estimator; the results in the remaining columns allow for cross-section dependence using either $\hat{\mu}_t^\bullet$ or cross-section averages in the FMOLS country regressions. In all cases the estimates presented are the unweighted means of the FMOLS country estimates. Intercept estimates as well as average estimates on cross-section averages in [4] and [5] of both panels are omitted to save space. Values in brackets are absolute standard errors following Pesaran and Smith (1995). Panel-t statistics are computed as $N^{-1/2}\sum_i t_i$ where t_i is the country-specific t-ratio for the estimate from the FM-OLS model. Panel B uses observations from only those countries for which variables were determined to be nonstationary (via country-specific ADF and KPSS testing). All models estimated in RATS.

Table 6: Country rankings by TFP-level

	Absolute Rank Difference between Implementations					
	[1]	[2]	[3]	[4]	[5]	[6]
	AMG–FE	**CMG–FE**	**Levels–FE**	**Levels–AMG**	**Levels–CMG**	**CMG–AMG**
Min	0.0	0.0	0.0	0.0	0.0	0.0
Mean	10.5	10.2	7.0	5.8	5.4	1.1
Median	10.0	10.0	5.0	5.0	4.5	1.0
IQR	7.5	9.0	7.0	7.0	6.8	1.0
Max	33.0	34.0	24.0	17.0	19.0	7.0

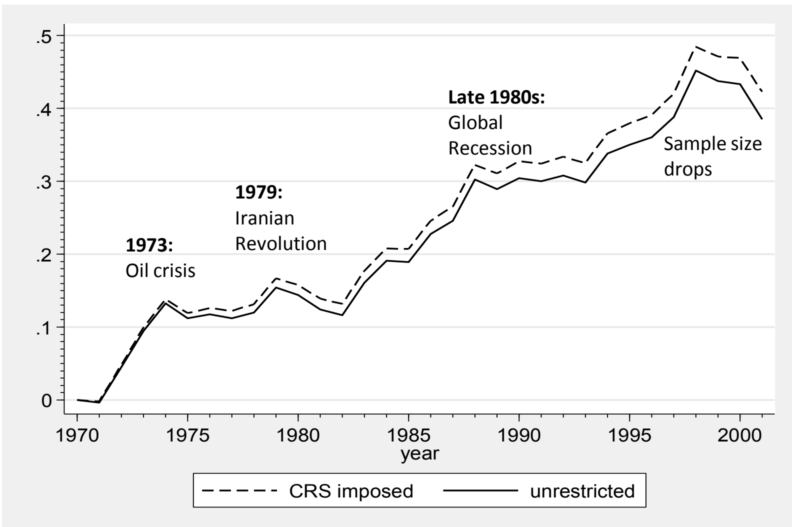
Notes: The table provides distributional statistics on the relative TFP level ranking (by magnitude) derived from the three regression models as well as the levels accounting for 1990: 'AMG–FE' is based on the absolute difference between TFP level rankings implied by the AMG and FE estimators, similarly for the other comparisons. FE refers to the Two-way Fixed Effects estimator, Table 3, Panel (A), column [2]; AMG refers to the Augmented Mean Group estimator, Table 3, Panel (B), column [3]. CMG refers to the Mean Group version of the Pesaran (2006) CCE estimator, ibid. column [5]. IQR reports the interquartile range of rank differences.

Figure 1: Technology Heterogeneity and Unobserved Common Factors

Factor loadings $\boldsymbol{\lambda}$		homogeneous	heterogeneous	
Factors $\boldsymbol{f}$		unrestricted	linear	unrestricted
Technology $\boldsymbol{\beta}$	homogeneous	POLS, 2FE, FD	FE w/ trends	CCEP, IFE
		$\alpha_{\{i\}} + \lambda f_t$	$\alpha_i + \lambda_i t$	$\alpha_i + \lambda_i' f_t$
		MRW, Islam, (CEL)	*MM*	*(CD)*
	heterogeneous	CD-MG	MG, GM-FMOLS	AMG, CMG
		$\alpha_i + \lambda f_t$	$\alpha_i + \lambda_i t$	$\alpha_i + \lambda_i' f_t$
			DKM, Pedroni	*ET, EHS*

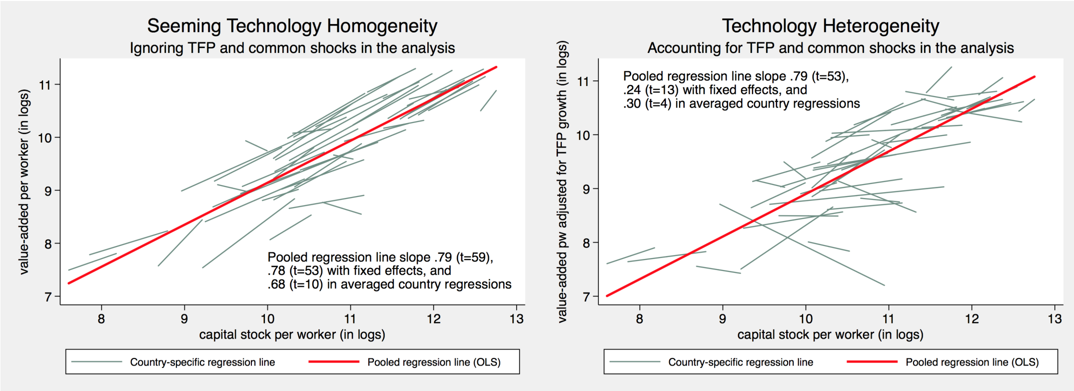
Notes: In addition to the various estimators we provide examples of empirical applications in the cross-country growth literature which adopted these implementations. MRW – Mankiw et al. (1992); Islam – Islam (1995); CEL – Caselli et al. (1996); MM – Martin and Mitra (2002); CD – Costantini and Destefanis (2009); DKM – Durlauf et al. (2001); Pedroni – Pedroni (2007); ET – Eberhardt and Teal (2013a); EHS – Eberhardt et al. (2013). A number of these references are in parentheses: Caselli et al. (1996) use the Arellano and Bond (1991) estimator while Costantini and Destefanis (2009) adopt the Bai et al. (2009) estimator, however their empirical specifications nevertheless fit into the respective cells in our schematic presentation. For each case we report the algebraic representation of how TFP is modelled when using this estimator. α referes to TFP levels, and the combination of λ and f_t (potentially nonlinear) or t (linear) to TFP evolution over time — refer to equation (1) for the encompassing model.

Figure 2: Evolution of 'average' TFP



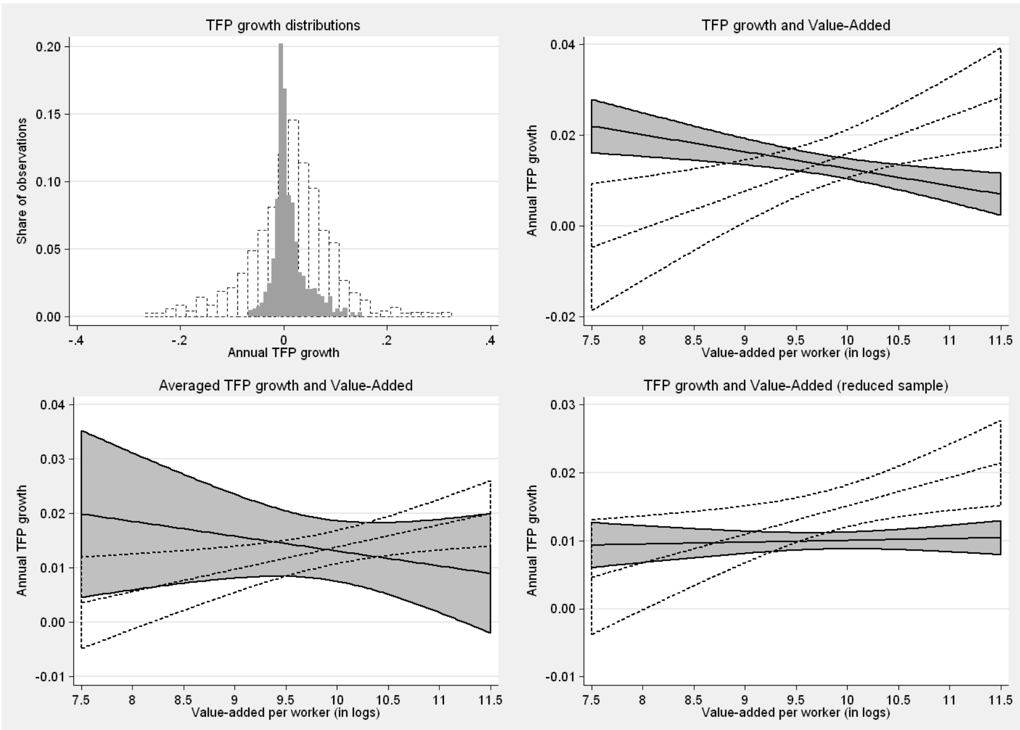
Notes: Derived from results in column [4], Panel (A) of Table 3.

Figure 3: Technology Heterogeneity in the Analysis of Development and TFP



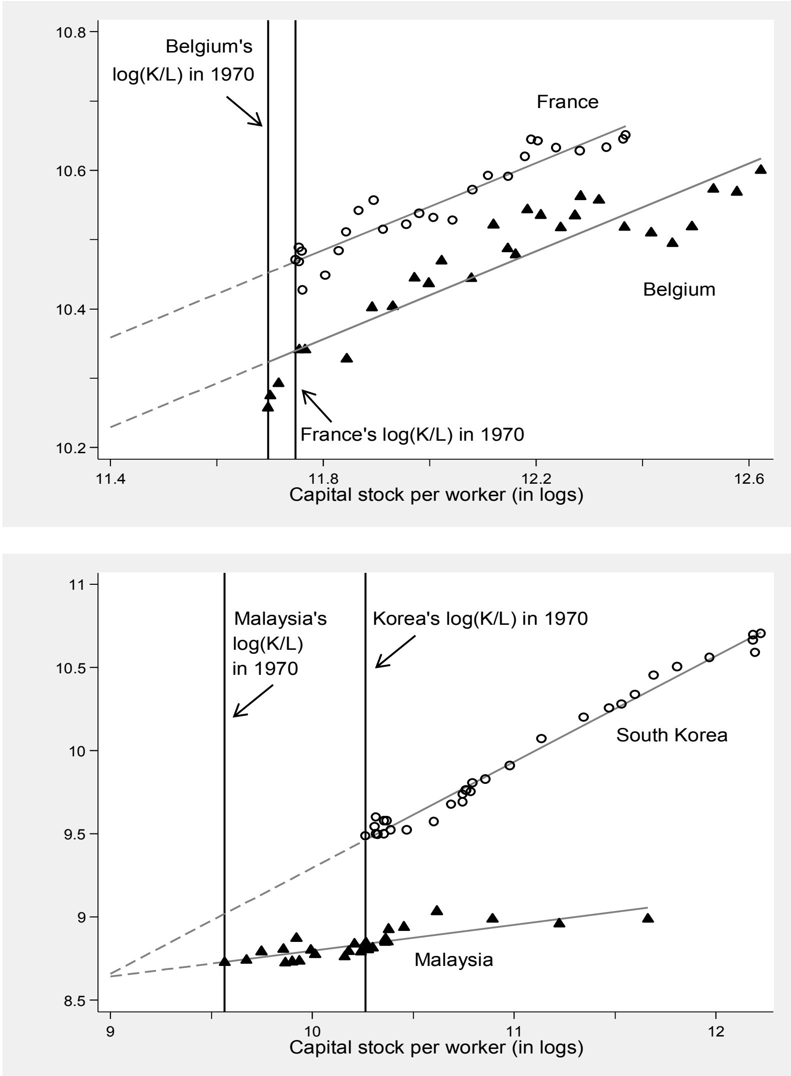
Notes: The graph on the left simple fits a linear regression line from country-specific data on manufacturing value-added per worker on manufacturing capital stock per worker (in logs) separately in 48 countries (teal-coloured lines), thus ignoring TFP evolution, spillovers and common shocks. The red line represents the pooled OLS regression slope. We further report the estimated slopes for pooled model with and without fixed effects and the mean slope for a naive Mean Group model (with country intercepts only). Values in parentheses are t-ratios. The graph on the right adjusts value-added per worker for annual country-specific TFP and the plots the same relationship across 48 countries. Although virtually identical, the pooled regression line in red here is for adjusted value-added per worker. Again we report slope coefficients for pooled OLS with and without country fixed effects, and the Mean Group result (which is a graphical representation of the AMG result in Panel B, column [3] of Table 3).

Figure 4: TFP growth from regression and growth accounting ($\beta^K = .33$)



Notes: We compare the TFP growth estimates derived from our preferred regression model, the AMG estimator, Table 3, Panel (B), column [3] (grey histogram and 90% confidence interval; capital coefficients differ across countries), with those obtained from simple TFP growth accounting (transparent histogram and dashed 90% confidence intervals; common capital coefficient: .33). Clockwise from the top left the graphs provide (i) histograms for these two sets of estimates, (ii) linear regression lines (and 90% confidence intervals) of TFP growth against log value-added per worker, (iii) as in (ii) but removing the top and bottom 5% of TFP growth estimates as computed in either exercise, and (iv) as in (ii) but using 48 country TFP growth averages.

Figure 5: Regression intercepts and TFP level estimates



Notes: In-sample (solid) and out-of-sample (dashed) linear prediction of the relationship between TFP-adjusted value-added per worker (on the y-axis) and capital stock per worker (on the x-axis), all variables in logarithms — see maintext for details on the TFP adjustment.

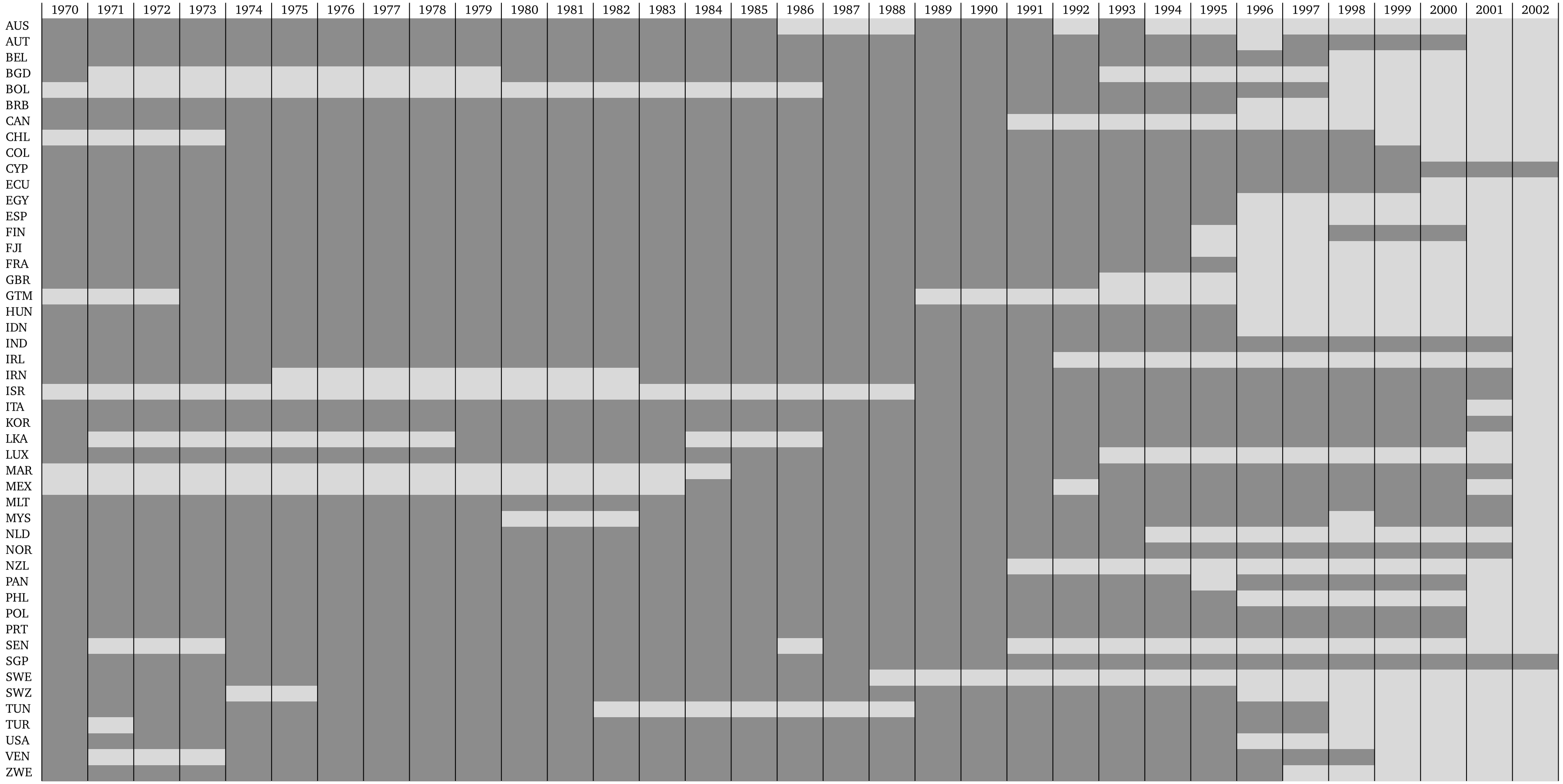
Technical Appendix

A Data

Data for output, value-added, material inputs and investment in manufacturing, all in current local currency units (LCU), are taken from the UNIDO Industrial Statistics 2004 (UNIDO, 2004), where material inputs were derived as the difference between output and value-added. The labour data series is taken from the same source, which covers 1963-2002. The capital stocks are calculated from investment data which has been transformed into constant US$ following the 'perpetual inventory' method (Klenow and Rodriguez-Clare, 1997). In order to make data in monetary values internationally comparable, it is necessary to transform all values into a common unit of analysis. We follow the transformations suggested by Martin and Mitra (2002) and derive all values in 1990 US$, using current LCU and exchange rate data from UNIDO, and GDP deflators from the UN Common Statistics database (UN, 2005), for which data are available from 1970-2003. Since our model is for a small open economy, we prefer using a single market exchange rates (LCU-US$ exchange rate for 1990) to purchasing-power-parity (PPP) adjusted exchange rates, since the latter are more appropriate when non-traded services need to be accounted. The resulting panel is unbalanced and has gaps within individual country time-series — see Figure TA-1. We have a total of $n = 1,194$ observations from $N = 48$ countries (min $T = 11$, max $T = 33$, average $T = 24$). We do not carry out any interpolation to fill these gaps and do not account for missing observations in any way. The preferred empirical specifications presented in the main section of our paper are based on heterogeneous parameter models, where arguably the unbalancedness (around 25% of observations in the balanced panel are missing) comes less to bear on the estimation results than in the homogeneous models due to the averaging of estimates. As a robustness check we also produced a 'cleaned' dataset where we applied mechanical 'cleaning rules' in order to address the most serious issues of measurement error,[1] which created a sample of $n = 872$ observations for $N = 38$ countries. The empirical results for this sample are virtually the same to those from the larger sample (available on request).

[1] We used the capital-to-materials ratio (K/M) to define a rule, bounded as $0.02 < K/M < 2$, and then dropped countries for which we had less than ten observations.

Figure TA-1: Missing observations



Notes: We indicate the data availability in our sample for each country over the 1970-2002 time horizon, where lighter shading signifies the observation as missing. In total 25% of observations are missing compared with a balanced panel.

B The Common Correlated Effects and Augmented Mean Group Estimators

The CCE estimators, developed by Pesaran (2006) and extended to nonstationary processes in Kapetanios, Pesaran and Yamagata (2011), augment the regression equation with cross-section averages of the dependent ($\bar{y}_t$) and independent variables ($\bar{\boldsymbol{x}}_t$) to account for the presence of unobserved common factors with heterogeneous impact.[2] For the Mean Group version (CMG), the individual country regression is specified as

$$y_{it} = a_i + \boldsymbol{b}_i' \boldsymbol{x}_{it} + c_{0i} \bar{y}_t + \sum_{m=1}^{k} \boldsymbol{c}_{mi} \bar{x}_{mt} + e_{it} \qquad \text{(TA-1)}$$

whereupon the parameter estimates $\hat{\boldsymbol{b}}_i$ are averaged across countries akin to the Pesaran and Smith (1995) Mean Group estimator.[3] The pooled version (CCEP) is specified as

$$y_{it} = a_i + \boldsymbol{b}' \boldsymbol{x}_{it} + \sum_{j=1}^{N} c_{0i} (\bar{y}_t D_j) + \sum_{m=1}^{k} \sum_{j=1}^{N} \boldsymbol{c}_{mi} (\bar{x}_{mt} D_j) + e_{it} \qquad \text{(TA-2)}$$

where the D_j represent country dummies.[4] The former is thus a simple extension to the Pesaran and Smith (1995) MG estimator based on on country-specific OLS regressions, whereas the latter is a standard fixed effects estimator augmented with additional regression terms.

In order to get an insight into the workings of this approach, consider the cross-section average of our model in equation (**??**): as the cross-section dimension N increases, given $\bar{\varepsilon}_t = 0$, we get

$$\bar{y}_t = \bar{\alpha} + \bar{\boldsymbol{\beta}}' \bar{\boldsymbol{x}}_t + \bar{\boldsymbol{\lambda}}' \boldsymbol{f}_t \qquad \Leftrightarrow \qquad \boldsymbol{f}_t = \bar{\lambda}^{-1} (\bar{y}_t - \bar{\alpha} - \bar{\boldsymbol{\beta}}' \bar{\boldsymbol{x}}_t) \qquad \text{(TA-3)}$$

This simple derivation provides a powerful insight: working with cross-sectional means of y and $\boldsymbol{x}$ can account for the impact of unobserved common factors (TFP) in the production process.[5] Given the assumed heterogeneity in the impact of unobserved factors across countries

[2] Parts of the discussion in this section is taken from Eberhardt and Teal (2013).

[3] Although $\bar{y}_t$ and e_{it} are not independent their correlation goes to zero as N becomes large.

[4] Thus in the MG version we have N individual country regressions with $2k+2$ RHS variables and in the pooled version a single regression equation with $k+(k+2)N$ RHS variables.

[5] Most conservatively the CCE estimators require $\bar{\boldsymbol{\lambda}} \neq 0$, i.e. that the impact of each factor is on average non-zero (Coakley, Fuertes and Smith, 2006). Alternative scenarios (see Pesaran, 2006; Kapetanios et al., 2011) allow for this assumption to be dropped in certain situations but for the sake of generality we maintain it here.

($\boldsymbol{\lambda}_i$) the estimator is implemented in the fashion detailed above, which allows for each country i to have different parameter estimates on $\bar{y}_t$ and the $\bar{\boldsymbol{x}}_t$, and thus implicitly on $\boldsymbol{f}_t$. Simulation studies (Pesaran, 2006; Coakley et al., 2006; Kapetanios et al., 2011; Pesaran and Tosetti, 2011) have shown that this approach performs well even when the cross-section dimension N is small, when variables are nonstationary, cointegrated or not, subject to structural breaks and/or in the presence of local spillovers and global/local business cycles. In the present study we implement two versions of the CCE estimators in the sector-level regressions: a standard form as described above; and a variant which includes the cross-section averages of the input and output variables in the own *as well as* the other sector. The latter specification allows for cross-section dependence *across* sectors, albeit at the cost of a reduction in degrees of freedom. It is conceivable that the evolution of the agricultural sector of developing countries influences that of the wider economy in general and the manufacturing sector in particular, such that this extension is sensible in the dual economy context.

Thus the Pesaran (2006) CCE estimators account for the presence of unobserved common factors by including cross-section averages of the dependent and independent variables in the regression equation and the estimates are obtained as averages of the individual country estimates, following the Pesaran and Smith (1995) MG approach. A related approach which we term the Augmented Mean Group (AMG) estimator (see Bond and Eberhardt, 2013, for details) accounts for cross-section dependence by inclusion of a 'common dynamic process' in the country regression. This process is extracted from the year dummy coefficients of a pooled regression in first differences (FD-OLS) and represents the levels-equivalent mean evolution of unobserved common factors across all countries. Provided the unobserved common factors form part of the country-specific cointegrating relation (Pedroni, 2007), the augmented country regression model encompasses the cointegrating relationship, which is allowed to differ across i.

$$\text{Stage (i)} \qquad \Delta y_{it} = \boldsymbol{b}'\Delta\boldsymbol{x}_{it} + \sum_{t=2}^{T} c_t \Delta D_t + e_{it} \quad \Rightarrow \hat{c}_t \equiv \hat{\mu}_t^{\bullet} \tag{TA-4}$$

$$\text{Stage (ii)} \qquad y_{it} = a_i + \boldsymbol{b}_i'\boldsymbol{x}_{it} + c_i t + d_i \hat{\mu}_t^{\bullet} + e_{it} \quad \hat{\boldsymbol{b}}_{AMG} = N^{-1}\sum_i \hat{\boldsymbol{b}}_i \tag{TA-5}$$

Stage (i) represents a standard FD-OLS regression with $T-1$ year dummies in first differences, from which we collect the year dummy coefficients (labelled as $\hat{\mu}_t^{\bullet}$). This process is extracted from the pooled regression *in first differences* since nonstationary variables and unobservables are believed to bias the estimates in the pooled *levels* regressions. In stage (ii) $\hat{\mu}_t^{\bullet}$ is included in each of the N standard country regressions which also include a linear trend term to capture omitted idiosyncratic processes evolving in a linear fashion over time. Alternatively we can subtract $\hat{\mu}_t^{\bullet}$ from the dependent variable, which implies the common process is imposed on each country with unit coefficient. In either case country-specific estimates are averaged across countries following the MG approach. Based on the results of Monte Carlo simulations (Bond and Eberhardt, 2013) we posit that the inclusion of $\hat{\mu}_t^{\bullet}$ allows for the *separate* identification of $\boldsymbol{\beta}_i$ or $\mathbb{E}[\boldsymbol{\beta}_i]$ and the unobserved common factors driving output and inputs, like in the CCE approach. In analogy, we can use $\Delta\hat{\mu}_t^{\bullet}$ in the country equations in first differences and can augment the Swamy (1970) RCM estimator in a similar fashion to yield the Augmented Random Coefficient Model (ARCM) estimators in levels and first differences — results for the ARCM were very similar to those in the AMG and in the interest of space are therefore omitted in the empirical section. We also applied an alternative version of the estimator where the first stage allows for heterogeneous slopes across countries: results for the AMG second stage are next to identical to those presented in Table **??**.

The focus of the CCE estimators is the estimation of consistent $\hat{\boldsymbol{b}}$ and not the nature of the unobserved common factors or their factor loadings: we cannot obtain an explicit estimate for the unobserved factors $\boldsymbol{f}_t$ or the factor loadings $\boldsymbol{\lambda}_i$, since the *average* impact of the factors ($\bar{\boldsymbol{\lambda}}$) is unknown. Our augmented estimators use an *explicit* rather than implicit estimate for $\boldsymbol{f}_t$ from the pooled first stage regression. Compared with the CCE approach we can obtain a simple but economically meaningful construct from the AMG setup: the common dynamic process $\hat{\mu}_t^{\bullet} = h(\bar{\boldsymbol{\lambda}}\boldsymbol{f}_t)$ represents common TFP evolution over time, whereby *common* is defined either in the literal sense, or as the sample mean of country-specific TFP evolution. The country-specific coefficient on the common dynamic process, $\hat{d}_i$ from equation (TA-5), represents the implicit factor loading on common TFP.

Immediate concerns about this augmented estimator relate to the issue of second stage 'regressions with generated regressors' (Pagan, 1984). However, simulation results (Bond and Eberhardt, 2013) suggest that the average standard error of the AMG estimates is of similar magnitude to the empirical standard deviation. A theoretical explanation is provided in Bai and Ng (2008), who show that second stage standard errors need not be adjusted for first stage estimation uncertainty if $\sqrt{T}/N \to 0$, as is arguably the case here.

C Investigating Cross-Section Correlation Properties

Table TA-1: Pesaran (2015) CD test

Variable	**CD-test**	***p*-value**	**corr**	**abs(corr)**
logs				
Value-added pw	68.15	0.000	0.447	0.627
Capital pw	81.12	0.000	0.530	0.639
Labour	16.72	0.000	0.115	0.629
annual growth rates				
Value-added pw	6.86	0.000	0.050	0.218
Capital pw	12.26	0.000	0.085	0.214
Labour	17.69	0.000	0.123	0.214

Notes: We implement the Pesaran (2015) test for null hypothesis of weak cross-section dependence for our main regression variables — in the upper panel the variables are in logs, in the lower panel in first differences of logs (growth rates). Due to the unbalanced nature of the panel it is not possible to obtain correlation coefficients for all 48 countries — in the tests for log variables we are forced to drop two countries (BOL, $n = 11$; ISR, $n = 13$) and in the growth rate versions three countries (BOL, $n = 10$; ISR, $n = 12$; MAR, $n = 16$) in order to make the test feasible. Under the null the test statistic is normally distributed. 'corr' and 'abs(corr)' report the average and average absolute correlation coefficients for each variable.

D Investigating Time Series Properties

Table TA-2: Second generation panel unit root tests

Pesaran (2007) panel unit root tests — CIPS♯														
output			**value-added**			**labour**			**capital**			**materials**		
lags	Z[t-bar]	(p)	lags	Z[t-bar]	(p)	lags	Z[t-bar]	(p)	lags	Z[t-bar]	(p)	lags	Z[t-bar]	(p)
0	-1.22	(.11)	0	-1.85	(.03)	0	2.39	(.99)	0	5.11	(1.00)	0	0.29	(.62)
1	0.01	(.51)	1	0.06	(.52)	1	1.26	(.90)	1	3.79	(1.00)	1	0.89	(.81)
1.42	1.13	(.87)	1.96	3.54	(1.00)	1.48	3.74	(1.00)	1.50	4.55	(1.00)	1.65	3.68	(1.00)
2	2.65	(1.00)	2	2.30	(.99)	2	4.21	(1.00)	2	3.96	(1.00)	2	1.05	(.85)
3	7.04	(1.00)	3	3.59	(1.00)	3	4.76	(1.00)	3	7.64	(1.00)	3	4.21	(1.00)
output/worker			**VA/worker**						**capital/worker**			**materials/worker**		
lags	Z[t-bar]	(p)	lags	Z[t-bar]	(p)				lags	Z[t-bar]	(p)	lags	Z[t-bar]	(p)
0	-1.08	(.14)	0	-2.55	(.01)				0	1.92	(.97)	0	0.57	(.72)
1	2.91	(1.00)	1	-0.73	(.23)				1	1.33	(.91)	1	3.74	(1.00)
1.44	5.98	(1.00)	1.65	3.77	(1.00)				1.71	5.92	(1.00)	1.83	9.62	(1.00)
2	5.02	(1.00)	2	2.37	(.99)				2	4.60	(1.00)	2	5.96	(1.00)
3	8.73	(1.00)	3	5.48	(1.00)				3	7.34	(1.00)	3	8.08	(1.00)

Notes: The CIPS test maintains the H_0 of a unit root process; augmentation with lags as indicated. ♯ All variables are in logs. In the third row for each variable in the lower panel we present the CIPS test statistic for 'ideal' lag augmentation of the underlying ADF regression (based on Akaike information criteria); the value for lags reported here is the *average* across countries.

E Diagnostic testing and robustness checks

We first investigated the density estimates for country-specific technology parameters estimated in the levels regressions using standard kernel methods with automatic bandwidth selection. The plots indicate that the distribution of these parameter estimates is symmetric around their respective means and roughly Gaussian, such that no significant outliers drive our results. We further carried out a number of residual diagnostic tests other than the analysis of stationarity and cross-section dependence. A cautious conclusion from these procedures would be that we are more confident about the country regression residuals possessing desirable properties (normality, homoskedasticity) than we are for their pooled counterparts (all results available on request).

Table TA-3: Gengenbach et al. (2016) cointegration tests

ECM-BASED COINTEGRATION TEST

no intercept	**AIC**		**BIC**		**10%**	**5%**	**1%**
$\bar{T}^*_{\alpha_y}$ (truncated)	-2.58	*	-2.75	**	-2.48	-2.55	-2.67
$\bar{T}^*_{\theta}$ (truncated)	25.66	**	25.61	**	12.10	12.43	13.07
avg. lag length	2.0		1.7				
intercept	**AIC**		**BIC**		**10%**	**5%**	**1%**
$\bar{T}^*_{\alpha_y}$ (truncated)	-2.63		-2.78		-2.86	-2.92	-3.03
$\bar{T}^*_{\theta}$ (truncated)	17.04	**	17.12	**	14.08	14.42	15.04
avg. lag length	2.3		1.7				
intercept, trend	**AIC**		**BIC**		**10%**	**5%**	**1%**
$\bar{T}^*_{\alpha_y}$ (truncated)	-2.44		-2.61		-3.23	-3.28	-3.39
$\bar{T}^*_{\theta}$ (truncated)	12.54		13.13		16.23	16.59	17.31
avg. lag length	2.1		1.8				

Notes: The $\bar{T}^*_{\alpha_y}$ and $\bar{T}^*_{\theta}$ statistics are averages of the N t-ratios and F-statistics from the country ECM regressions, where extreme t-ratios/F-statistics have been replaced by bounds (truncated; we used $\varepsilon = .000001$) following the strategy devised in Gengenbach et al. (2016). This paper also provides simulated critical values we present here ($N = 50$) in the supplementary material. Both test statistics are one-sided: for the $\bar{T}^*_{\alpha_y}$ large negative values lead to rejection of the null, whereas for the $\bar{T}^*_{\theta}$ it is large positive values which lead to rejection. H_0 in all cases: no error correction, i.e. no cointegration; lag-length p_i determined using AIC or BIC as indicated; $p_x = 2$ (capital per worker and $\hat{\mu}^\bullet_t$).

Cointegration tests are commonly carried out as a *pre*-estimation testing procedure, however we have delayed these until *after* estimation since we hypothesise that unobservable TFP forms part of the cointegrating vector. Employing our first stage estimate $\hat{\mu}^\bullet_t$ we carried out a cointegration testing procedure based on the error correction model representation, first introduced by Westerlund (2007) and refined by Gengenbach et al. (2016). Results in Table TA-3 imply that there are good grounds to suggest that value-added per worker, capital per worker and our estimate for TFP are heterogeneously cointegrated.

F Parameter heterogeneity tests

The individual country coefficients emerging from our regressions imply considerable parameter heterogeneity across countries. However, this apparent heterogeneity may be due to sampling variation and the relatively limited number of time-series observations in each country individually (Pedroni, 2007). We therefore carry out a number of parameter heterogeneity tests for the results from the various CMG and augmented MG/RCM estimations.

As a *first test*, we compute the residuals in the case of parameter homogeneity for each country

$$\begin{aligned} H_{het} &\equiv o^{\bullet}_{it} - \bar{b}\, k_{it} - \bar{c}\, m_{it} - \bar{\mu} t - \bar{A}_0 \\ H_{het} &\equiv o_{it} - \bar{b}\, k_{it} - \bar{c}\, m_{it} - \bar{\mu} t - \bar{d}\hat{\mu}^{\bullet}_t - \bar{A}_0 \end{aligned}$$

where $\bar{b}$, $\bar{c}$ and $\bar{d}$ (for Augmented models) are the *mean* estimates for capital per worker (k), materials per worker (m) and the common dynamic process taken from the results in Table **??** in the paper, with $\bar{\mu}$ the average country trend term and $\bar{A}_0$ the average intercept term (the latter is not important for this analysis). The common dynamic process is either subtracted from the output variable ($o^{\bullet}_{it}$) or included as indicated above. Similarly for the other models, the VA specifications and the specifications in first differences. In a second step, we regress the residuals created on the input variables, a country trend or drift term and country- and year-dummies in a pooled regression

$$H_{het} = \pi_b\, k_{it} + \pi_c\, m_{it} + \pi_d\, t\, (+ \sum_i \pi_{e,i}) \qquad \text{(TA-6)}$$

The rationale behind this test is as follows: if factor input parameters were truly heterogeneous across countries, we would expect the pooled regression to produce statistically significant coefficients ($\pi_j \neq 0$). Results are presented in Tables TA-4 and TA-5.

As can be seen the levels regressions imply that capital parameter homogeneity is rejected, while the materials coefficients are more likely to be homogeneous (in the gross output specification). In the VA-specification capital parameter homogeneity is rejected in all models. In contrast the tests for the first difference specifications on the whole do not provide much evidence for heterogeneity, with all covariates insignificant with the exception of the case of CMG in first differences. Note that the kernel densities for the technology parameters underlying the above heterogeneity tests do not differ considerably between levels and FD specifications (FD densities not reported). This stark difference is therefore likely to be driven by the impact of nonstationarity on the test.

Table TA-4: Parameter Heterogeneity — Pooled Tests (levels)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
estimator	**MG**	**RCM**	**AMG(i)**	**AMG(ii)**	**CMG(i)**	**CMG(ii)**	**ARCM(i)**	**ARCM(ii)**
dependent variable$^\sharp$	H_{het}	H_{het}	H_{het}^{va}	H_{het}^{va}	H_{het}	H_{het}^{va}	H_{het}^{va}	H_{het}
regressors								
capital pw	0.4704	0.4242	0.3733	0.3511	0.197	0.3517	0.3072	0.3083
	[15.94]**	[14.38]**	[12.95]**	[11.90]**	[6.55]**	[12.00]**	[10.65]**	[10.68]**
country trends	-0.0112	-0.0026	-0.0088	0.0039	-0.0018	-0.0087	-0.0068	-0.0071
	[11.72]**	[2.73]**	[9.07]**	[4.07]**	[1.87]**	[9.13]**	[7.04]**	[7.33]**
intercept terms	all sign	all sign	all sign	all sign	all sign	all sign	all sign	all sign
	at 1%	at 1%	at 1%	at 1%	at 1%	at 1%	at 1%	at 1%
obs	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194

Notes: All variables are in logs. Values in brackets are absolute t-statistics. The models underlying the construction of H_{het} are presented in Table **??** in the main text. We indicate statistical significance at the 5% and 1% level by * and ** respectively.

Table TA-5: Parameter Heterogeneity — Pooled Tests (FD)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
estimator	**ΔMG**	**ΔRCM**	**ΔAMG(i)**	**ΔAMG(ii)**	**ΔCMG(i)**	**ΔCMG(ii)**	**ΔARCM(i)**	**ΔARCM(ii)**
dependent variable$^\sharp$	H_{het}	H_{het}	H_{het}	H_{het}	H_{het}	H_{het}	H_{het}	H_{het}
regressors								
capital pw	0.0989	0.0546	0.0157	0.0069	-0.0791	-0.0202	-0.0162	-0.0106
	[1.11]	[0.61]	[0.18]	[0.08]	[0.88]	[0.22]	[0.18]	[0.12]
drift terms	only 2	only 2	only 2	only 2	only 2	only 2	only 2	only 2
	sign.	sign.	sign.	sign.	sign.	sign.	sign.	sign.
obs	1,128	1,128	1,128	1,128	1,128	1,128	1,128	1,128

Notes: See Table TA-4 for details. The results for the country regressions in first difference tested here for parameter heterogeneity are presented in Table **??** in the main text.

Secondly, we report the Swamy (1970) $\hat{S}$ statistic from the gross output and VA regressions in levels and first differences in Table TA-6.[6] For a detailed discussion of this test see Pesaran and Yamagata (2008). Note that the test for the equation in levels is testing heterogeneity of *all* parameters, including the intercepts; since the assumption of heterogeneous TFP levels is rather uncontroversial, this test does not adquately address our interest in the homogeneity of

[6]The levels and FD tests are taken from the regressions in Table **??** of the paper.

technology parameters. We therefore also provide a test for the levels specification where the intercept terms have been dispensed with via transformation of the data into mean-deviations. Estimates for this specification are of course identical to those of the untransformed levels equation.

Table TA-6: Parameter Heterogeneity — Swamy (1970) Tests

	RCM			**ARCM**		
Specification	(a)	(a)’	(b)	(b)’	(c)	(c)’
	Levels	MD	Levels	MD	Levels	MD
	51,123.0 (.00)	1,531.6 (.00)	62,499.9 (.00)	1,440.4 (.00)	69,157.0 (.00)	1726.7 (.00)
	FD		FD		FD	
	191.1 (.00)		153.5 (.00)		258.7 (.00)	

Notes: Swamy $\hat{S}$ is distributed χ^2 with $k(N-1)$ degrees of freedom. † Data in mean-deviations.

The Swamy $\hat{S}$ test rejects parameter heterogeneity for all specifications tested. In general, this test was developed for panels where N is large relative to T. Using Monte Carlo experiments, Pesaran and Yamagata (2008) show that in case of a panel of $T = 30, N = 50$ the test has power but tends to over-reject — a tendency which becomes worse with the number of parameters included in the model.[7] Further, as Pedroni (2007) points out, the Swamy-based tests are not designed for nonstationary panel data.

Thirdly, we produce Wald statistics, as suggested by Canning and Pedroni (2008)

$$W_\theta = \sum_i \frac{(\hat{\theta}_i - \bar{\theta})^2}{\mathbb{V}ar(\theta_i)} \qquad\qquad W_\theta \sim \chi^2(N)$$

where $\hat{\theta}_i$ is the parameter coefficient from the country regression, $\bar{\theta}$ is the unweighted average parameter estimate and $\mathbb{V}ar(\theta_i)$ its variance across all countries. If parameters are similar across countries, the test statistic will be small, whereas if parameters are heterogeneous W_θ will be larger. The validity of this test depends on T being moderate to large. The null for this test is that *all* countries have the same parameter value. Table TA-7 presents the summed Wald statistics for the entire sample, as well as an indication of the share of country-specific tests rejecting the null of equality between country estimate and full sample mean estimate (for both the levels and FD specifications).

The Wald tests reject homogeneity for the factor parameters derived from the levels models

[7] The adjusted Swamy statistic $\tilde{\Delta}_{adj}$ developed by the same authors, although appropriately sized, suffers from low power in a sample of $T = 30, N = 50$, in particular if errors are non-normal.

in case of both the gross-output and value-added specifications. The statistics are particularly large for the trend terms in the levels specifications, thus rejecting homogeneity emphatically, which is not always the case for the drift terms in the first-difference specifications. Turning to the share of countries rejecting parameter homogeneity, it can be seen that roughly half of all countries reject homogeneity for all covariates in the levels specifications. This share falls to less than one third in the models in first differences.

Fourthly, following Pedroni (2007), we produce an F-statistic for the standard and augmented MG and RCM regression models (Pesaran and Yamagata, 2008, p.52)

$$\begin{aligned} F &= \left(\frac{RSS_{hom} - RSS_{het}}{RSS_{het}}\right)\left(\frac{df_D}{df_N}\right) \\ F &\sim F(df_N, df_D) \qquad df_N = k \times (N-1) \qquad df_D = N(\bar{T} - k - 1) \end{aligned}$$

where k is the number of parameters in each country-regression and RSS_{hom} and RSS_{het} are the sums of the squared residuals of the homogeneous and heterogeneous models respectively — in the former case the mean coefficient estimates are imposed. This tests the full parameter heterogeneity versus the full homogeneity case. We do not compute F-tests for the CMG models, as the parameters on the period-average are not meant to be identical.

The F tests are valid for fixed N, when the regressors are strictly exogenous and the error variances are homoskedastic (Pesaran and Yamagata, 2008).[8] All of the test results presented in Table TA-8 reject parameter homogeneity for the factor input variables at the 1% level of significance. It is intuitive why the test statistics may emphatically reject the null: if the homogeneity restriction is incorrect, the country regressions do not cointegrate under the null, such that the regression errors will be nonstationary. As a result the F-statistic will quickly diverge and reject the null (Pedroni, 2007).

Like in the Swamy $\hat{S}$ Test we are faced with the problem that the tests evaluate the *full* regression model for the null of parameter homogeneity, which is not sensible in the levels regression case since heterogeneous intercepts are commonly accepted in the literature. In order to by-

[8]In the levels specifications, $k = 4$ includes technology parameters, intercept and trend terms ($k = 3$ in the VA case); in the first difference ones $k = 3$ includes technology parameters and drift terms ($k = 2$ in the value-added specification). N is the number of countries, 48, and $\bar{T}$ represents the average time series length, s.t. $N\bar{T} = 1,162$ (VA:$1,194$) in the levels and $N\bar{T} = 1,094$ ($1,128$) in the FD case.

Table TA-7: Parameter Heterogeneity — Wald Tests (levels and FD)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Specification in levels								
estimator	**MG**	**RCM**	**AMG(i)**	**AMG(ii)**	**CMG(i)**	**CMG(ii)**	**ARCM(i)**	**ARCM(ii)**
full sample W_θ								
capital pw (k)	578.1**	388.1**	474.0**	542.3**	515.5**	548.0**	320.1**	354.9**
country trends	1,044.3**	720.9**	781.8**	267.7**		374.4**	636.7**	194.5**
Country-specific $W_{\theta,i}$								
share rejecting H_0: k	52%	46%	52%	56%	54%	46%	50%	54%
share rejecting H_0: t	65%	56%	58%	44%		42%	52%	40%

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Specification in FD								
estimator	**MG**	**RCM**	**AMG(i)**	**AMG(ii)**	**CMG(i)**	**CMG(ii)**	**ARCM(i)**	**ARCM(ii)**
full sample W_θ								
capital pw (k)	166.1**	147.1**	117.1**	139.8**	125.2**	141.5**	115.9**	132.4**
country drifts	80.5**	73.5*	51.7	109.5**		74.7**	54.0	86.8**
Country-specific $W_{\theta,i}$								
share rejecting H_0: k	25%	33%	33%	31%	35%	29%	31%	33%
share rejecting H_0: drift	19%	17%	15%	21%		15%	15%	21%

Notes: Analysis for 1,194 observations (1,128 in the first difference specifications) in 48 countries. The models underlying the construction of the Wald statistics are presented in Table **??** in the main text. In the full sample tests $W_\theta \sim \chi^2(48)$, with 5% and 1% critical values 65.17 and 73.70 respectively ($W_\theta = \sum_i W_{\theta,i}$); for country-specific tests ($W_{\theta,i}$) we apply the 10% critical value of 2.7. The null hypothesis in all cases is parameter homogeneity. For W_θ we indicate statistical significance at the 5% and 1% level by * and ** respectively.

Table TA-8: Parameter Heterogeneity — F-Tests

	[1]	[2]	[3]	[4]	[5]	[6]
estimator	**MG**	**RCM**	**AMG(i)**	**AMG(ii)**	**ARCM(i)**	**ARCM(ii)**
levels						
F	413.7 (.00)	334.1 (.00)	339.9 (.00)	279.4 (.00)	287.5 (.00)	232.3 (.00)
distr	$F(141,1002)$	$F(141,1002)$	$F(141,1002)$	$F(188,954)$	$F(141,1002)$	$F(188,954)$
first differences						
F	2.0 (.00)	1.5 (.00)	2.6 (.00)	2.4 (.00)	1.6 (.00)	1.8 (.00)
distr	$F(94,950)$	$F(94,950)$	$F(94,950)$	$F(141,902)$	$F(94,950)$	$F(141,902)$

Notes: See above text for construction of the Panel F statistic. The models underlying the construction of the F statistics are presented in Table **??** in the main text. The null hypothesis in all cases is parameter homogeneity.

pass this problem we also computed F-statistic for the levels MG and Augmented MG cases where the intercepts have been dispensed with by taking all variables in deviations from the country-mean — all of these reject parameter homogeneity at the 1% level.

Taken together the various diagnostic tests we carried out in this section do give a strong indication that parameter homogeneity is rejected. The differences in the results for levels and first difference specifications however indicate that nonstationarity may drive some of the results reported. Nevertheless, even if heterogeneity were not very significant in qualitative terms, our contrasting of pooled and country regression results in the paper has shown that it nevertheless matters greatly for correct empirical analysis in the case of nonstationary variable series.

Further parameter heterogeneity tests were considered for this analysis: Pesaran and Yamagata (2008) compare their own version of Swamy's test of parameter homogeneity (denoted $\tilde{\Delta}$) with the 'traditional' Swamy test and F-Test we computed above, a Hausman-type comparison of Fixed Effects and Mean Group estimates and the Phillips and Sul (2003) G-test. Their Monte Carlo experiments suggest that all of these tests have low power in panels with the dimensions we observe ($N = 48, T \approx 24$) and we therefore did not further pursue any of these here.

G The growth accounting literature

Empirical studies using TFP growth accounting have a long tradition since Abramowitz (1956), Kendrick (1956) — who coined the term Total Factor Productivity — and Solow (1957). Under standard assumptions value-added growth is decomposed into contributions of inputs and TFP growth, imposing a common capital coefficient β^K

$$\Delta y_{it} = \beta^K \Delta k_{it} + \Delta \mathrm{TFP}_{it} \qquad \Leftrightarrow \qquad \Delta \mathrm{TFP}_{it} = \Delta y_{it} - \beta^K \Delta k_{it} \tag{TA-7}$$

The simple computation as well as function-free nature of this approach represent considerable strengths and in part explain its popularity. The accounted TFP growth is in theory disembodied, Hicks-neutral exogenous technical progress. In practice however, one needs to keep in mind that TFP is a residual, such that it represents a 'catch-all' for output growth that cannot be explained by factor accumulation. Thus if TFP growth is recovered via growth accounting its coefficient "need not represent only technological change and may not represent technological change at all" (Baier, Dwyer and Tamura, 2006, p.27) since measurement error, violations of assumptions and 'incorrect' variable construction can cause considerable bias. Any measurement error in output, labour or capital enters the residual term and thus TFP growth. This may have considerable impact since factor inputs need to account correctly for embodied technical change, which given the difficulty of distinguishing between embodied and disembodied technical progress seems impossible (Lipsey and Carlaw, 2001; Baier et al., 2006). The method further cannot disentangle the underlying endogeneity problem, such that inputs cannot be argued to *cause* output (Gollin, 2010). Violations of the assumptions of constant returns to scale, and private and social marginal product equality can add to further accounting error (Barro, 1999) — conceptually, the simple accounting framework for instance runs counter to the substantial empirical literature on knowledge spillovers (Eberhardt, Helmers and Strauss, 2013). As a result, it is now widely recognised that TFP growth derived from growth accounting "does not really measure technical change" (Caselli, 2008), nevertheless most empirical work takes findings of substantial TFP growth as a very positive and meaningful insight into the growth process.

Technical Appendix References

Abramowitz, Moses (1956). "Resource and output trends in the United States since 1870." *American Economic Review*, Vol. 46(2): 5–23.

Bai, Jushan and Ng, Serena (2008). "Large Dimensional Factor Analysis." *Foundations and Trends in Econometrics*, Vol. 3(2): 89–163.

Baier, Scott L., Dwyer, Gerald P. and Tamura, Robert (2006). "How Important are Capital and Total Factor Productivity for Economic Growth?" *Economic Inquiry*, Vol. 44(1): 23–49.

Barro, Robert J. (1999). "Notes on Growth Accounting." *Journal of Economic Growth*, Vol. 4(2): 119–137.

Bond, Stephen R. and Eberhardt, Markus (2013). "Accounting for unobserved heterogeneity in panel time series models." Unpublished mimeo, November.

Canning, David and Pedroni, Peter (2008). "Infrastructure, Long-Run Economic Growth And Causality Tests For Cointegrated Panels." *Manchester School*, Vol. 76(5): 504–527.

Caselli, Francesco (2008). "Growth Accounting." In: Steven N. Durlauf and Lawrence E. Blume (Editors), "The New Palgrave Dictionary of Economics Online," (Palgrave Macmillan), 2nd edition.

Coakley, Jerry, Fuertes, Ana-Maria and Smith, Ron P. (2006). "Unobserved heterogeneity in panel time series models." *Computational Statistics & Data Analysis*, Vol. 50(9): 2361–2380.

Eberhardt, Markus, Helmers, Christian and Strauss, Hubert (2013). "Do spillovers matter when estimating private returns to R&D?" *The Review of Economics and Statistics*, Vol. 95(2): 436–448.

Eberhardt, Markus and Teal, Francis (2013). "Structural Change and Cross-Country Growth Empirics." *World Bank Economic Review*, Vol. 27: 229–71.

Gengenbach, Christian, Westerlund, Joakim and Urbain, Jean-Pierre (2016). "Error Correction Testing in Panels with Common Stochastic Trends." *Journal of Applied Econometrics*, Vol. 31: 982–1004.

Gollin, Doug (2010). "Agricultural productivity and economic growth." In: P Pingali and R Evenson (Editors), "Handbook of Agricultural Economics," Vol. 18, chapter 71 (Elsevier North-Holland), pp. 3825–3866.

Kapetanios, George, Pesaran, M. Hashem and Yamagata, Takashi (2011). "Panels with Non-stationary Multifactor Error Structures." *Journal of Econometrics*, Vol. 160(2): 326–348.

Kendrick, John W. (1956). "Productivity Trends: Capital and Labor." *The Review of Economics and Statistics*, Vol. 38(3): 248–257.

Klenow, Peter J. and Rodriguez-Clare, Andres (1997). "Economic growth: A review essay." *Journal of Monetary Economics*, Vol. 40(3): 597–617.

Lipsey, Richard G. and Carlaw, Kenneth I. (2001). "What does Total Factor Productivity measure?" Unpublished manuscript.

Martin, Will and Mitra, Devashish (2002). "Productivity Growth and Convergence in Agriculture versus Manufacturing." *Economic Development and Cultural Change*, Vol. 49(2): 403–422.

Pagan, Adrian (1984). "Econometric Issues in the Analysis of Regressions with Generated Regressors." *International Economic Review*, Vol. 25(1): 221–47.

Pedroni, Peter (2007). "Social capital, barriers to production and capital shares: implications for the importance of parameter heterogeneity from a nonstationary panel approach." *Journal of Applied Econometrics*, Vol. 22(2): 429–451.

Pesaran, M. Hashem (2006). "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica*, Vol. 74(4): 967–1012.

Pesaran, M. Hashem and Smith, Ron P. (1995). "Estimating long-run relationships from dynamic heterogeneous panels." *Journal of Econometrics*, Vol. 68(1): 79–113.

Pesaran, M. Hashem and Tosetti, Elisa (2011). "Large Panels with Common Factors and Spatial Correlations." *Journal of Econometrics*, Vol. 161(2): 182–202.

Pesaran, M. Hashem and Yamagata, Takashi (2008). "Testing Slope Homogeneity In Large Panels." *Journal of Econometrics*, Vol. 142(1): 50–93.

Phillips, Peter and Sul, Donggyu (2003). "Dynamic panel estimation and homogeneity testing under cross section dependence." *Econometrics Journal*, Vol. 6(1): 217–259.

Solow, Robert. M. (1957). "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics*, Vol. 39(3): 312–20.

Swamy, P. A. V. B. (1970). "Efficient Inference in a Random Coefficient Regression Model." *Econometrica*, Vol. 38(2): 311–23.

UN (2005). "UN Common Statistics 2005." Online database, New York: UN.

UNIDO (2004). "UNIDO Industrial Statistics 2004." Online database, Vienna: UNIDO, united nations Industrial development organisation.

Westerlund, Joakim (2007). "Testing for Error Correction in Panel Data." *Oxford Bulletin of Economics and Statistics*, Vol. 69(6): 709–748.