University of Nottingham
UK | CHINA | MALAYSIA

research paper series

Globalisation, Productivity and Technology

Research Paper 2017/11

Measuring Productivity and Absorptive Capacity Evolution

By
Steff De Visscher, Markus Eberhardt and Gerdie Everaert

GEP

Measuring Productivity and Absorptive Capacity Evolution*

Stef De Visscher[1,2], Markus Eberhardt[3,4], and Gerdie Everaert[1]

[1] *Ghent University, Belgium*
[2] *Research Foundation Flanders (FWO), Belgium*
[3] *University of Nottingham, U.K.*
[4] *Centre for Economic Policy Research, U.K.*

September 15, 2017

Abstract: We develop a new way to estimate cross-country production functions which allows us to parametrize unobserved non-factor inputs (total factor productivity) as a global technology process combined with country-specific time-varying absorptive capacity. The advantage of our approach is that we do not need to adopt proxies for absorptive capacity such as investments in research and development (R&D) or human capital, or specify explicit channels through which global technology can transfer to individual countries, such as trade, foreign direct investment (FDI) or migration: we provide an endogenously-created index for relative absorptive capacity which is easy to interpret and encompasses potential proxies and channels. Our implementation adopts an unobserved component model and uses a Bayesian Markov Chain Monte Carlo (MCMC) algorithm to obtain posterior estimates for all model parameters. This contribution to empirical methodology allows researchers to employ widely-available data for factor inputs (capital, labor) and GDP or value-added in order to arrive at policy-relevant insights for industrial and innovation policy. Applying our methodology to a panel of 31 advanced economies we chart the dynamic evolution of global TFP and country-specific absorptive capacity and then demonstrate the close relationship between our estimates and salient indicators of growth-enhancing economic policy.

JEL Classifications: O33, F43, F60, C23, C21

Keywords: total factor productivity, absorptive capacity, common factor model, time-varying parameters, unobserved component model, MCMC

*We thank seminar and workshop participants for useful comments and suggestions. The usual disclaimers apply. Markus Eberhardt gratefully acknowledges funding from the U.K. Economic and Social Science Research Council [grant number ES/K008919/1]. Correspondence: Markus Eberhardt, School of Economics, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, United Kingdom. Email: markus.eberhardt@nottingham.ac.uk

1 Introduction

Output per capita shows enormous and persistent differences across countries. As variations in factor inputs are unable to explain these differences, there is an important role for disparities in Total Factor Productivity (TFP). The relative importance of TFP vis-à-vis factor accumulation for economic growth has occupied economists not least since Tinbergen (1942), Abramovitz (1956) and Solow (1956). One strand of the literature proceeded to give TFP a more structural interpretation, namely that of successfully assimilated global technology (Parente and Prescott, 1994, 2002). What unites concepts such as 'absorptive capacity' and alternatives – e.g. social capability (Abramovitz, 1986) – is the notion that despite the designation of knowledge as a public good or being in the public domain (Nelson, 1959; Arrow, 1962) technological catch-up is by no means guaranteed, but requires considerable efforts and investments (Aghion and Jaravel, 2015).[1]

In the empirical growth literature there is a long tradition of quantifying 'foreign' elements of TFP by assuming specific channels through which international knowledge spillovers can occur and/or pin-pointing country characteristics deemed synonymous with absorptive capacity. The most prominent channels are arguably the patterns of international trade, foreign direct investment and international migration/personal interaction (Coe and Helpmann, 1995; Pottelsberghe and Lichtenberg, 2001; Madsen, 2007, 2008; Acharya and Keller, 2009; Andersen and Dalgaard, 2011; He and Maskus, 2012, see Keller (2004, 2010) for detailed surveys). Human capital (Griffith et al., 2004; Madsen et al., 2010; Ertur and Musolesi, 2017) and investment in R&D (Cohen and Levinthal, 1989; Aghion and Howitt, 1998; Griffith et al., 2003, 2004) and their interactions are frequently employed as proxies for absorptive capacity.[2]

While *a priori* all of these factors and channels are likely to be relevant to capture the discovery, assimilation and exploitation of ideas and innovations developed elsewhere, our approach overcomes two major difficulties facing the empirical analysis of absorptive capacity: (i) the possible bias in estimates for the proxies included if these are correlated with other, omitted determinants (e.g. Acharya, 2016; Corrado et al., 2017). This criticism points to recent efforts to quantify productivity-enhancing expenditure in intangible capital, of which formal research and development is just one element. It can further be argued that the recent wave of research on the importance of management for productivity (see Bloom et al., 2016, for a recent overview) highlights the shortcomings of a 'narrow' R&D expenditure focus in the empirical work on knowledge diffusion and absorption to date. And (ii) the presence of cross-section correlation in the panel induced by either spillovers or common shocks, as highlighted in the case of private returns to R&D and knowledge spillovers in

[1] For a detailed discussion of the origins of absorptive capacity seeFagerberg et al. (2009). In this article we use knowledge spillovers synonymously with 'technology spillovers' or more broadly the assimilation of ideas and innovations developed in other countries. Technology is used interchangeably with productivity, knowledge and TFP.

[2] A recent theoretical paper by Strulik (2014) links historical knowledge diffusion to capital accumulation.

a recent paper by Eberhardt et al. (2013). These authors show that private returns to R&D are dramatically lower once knowledge spillovers and common shocks are taken into account. At the same time, they hint that the results in the existing empirical literature on knowledge spillovers following the seminal contribution by Coe and Helpmann (1995) are likely unreliable due to omitted variable bias induced by the presence of common shocks with heterogeneous impact.

In this paper we propose a novel way to analyse cross-country productivity which uses the cross-section dimension of the panel data. Rather than imposing explicit channels and/or relying on individual proxies for knowledge spillovers and absorption we parametrize the relationship between 'free' global knowledge and a country's (restricted) capacity to appropriate this knowledge by adding a common factor error structure to a log-linearised Cobb-Douglas specification for aggregate GDP with factor inputs labour and capital stock (Bai et al., 2009).[3] Our model is in the neoclassical tradition, in that we do not explain the determinants of global TFP, but also captures some of the defining features of second generation endogenous growth models and their empirical implementations, namely (i) that TFP evolution is not identical across countries, even in the long-run (Evans, 1997; Lee et al., 1997), and (ii) that TFP is not limited to the innovation efforts of individual countries, but is made up to a significant extent of spillovers of knowledge from elsewhere (Eaton and Kortum, 1994, 1999; Aghion and Howitt, 1998; Klenow and Rodríguez-Clare, 2005).

The estimated patterns in the country-specific evolution of absorptive capacity and global TFP we present below are of interest in their own right because they provide insights into the empirical validity of stylised theoretical models without relying on narrow proxies for absorptive capacity (such as human capital or R&D investment). Acharya (2016) highlights the role of 'intangible capital' in knowledge spillovers, providing a broader interpretation than R&D investments to include other proxies (ICT capital compensation to gross output) in his regressions. Haskel et al. (2013) and Corrado et al. (2017) have recently made concerted efforts to quantify non-R&D intangibles so as to capture organisational capital, skills and training, etc. at the macro level. The empirical implementation adopted in our study allows us to capture all aspects of knowledge, its international transmission channels and domestic absorptive efforts. Our methodological contribution extending Pesaran's (2006) Common Correlated Effects (CCE) estimator to include time-varying factor loadings further provides the building blocks to incorporate a much richer empirical specification to jointly determine the respective causal effects of trade, innovation effort, human capital, etc. on economic growth and inter-country knowledge spillovers.

We estimate our model using panel data for 31 advanced economies covering 1953-2014. Our first finding, based on a stochastic model specification search, is that there is relevant time variation in

[3] Factor models have gained popularity in empirical macroeconomics over the past decade in the analysis of business cycle (Miyamoto and Nguyen, 2017; Kose et al., 2012) and forecasting (Elliott and Timmermann, 2005) among other applications.

absorptive capacities. Especially countries like Ireland, South Korea and Taiwan have been able to increase their ability to assimilate foreign knowledge over the sample period. Second, for the vast majority of economies in our sample changes in absorptive capacity are *not* found to have permanent growth effects, but merely affect the level of TFP, in line with models presented in Klenow and Rodríguez-Clare (2005). Our third finding is that the endogenously-created estimate for global TFP evolution shows substantial decline over time, from a stable 2% per annum up to the late 1960s to 1% per annum since the mid-1990s. This evolution is primarily the outcome of a substantial decline in TFP growth during the 1970s and 1980s. With respect to most recent developments in the wake of the Global Financial Crisis it is too early to establish whether TFP growth will stabilise at the pre-crisis mean, has entered into a new period of secular decline or experienced a structural break to a new, lower level close to zero. Our fourth finding relates to the country-specific evolution of absorptive capacity, which in tandem with the estimate for base year can be squared with existing information on the timing and extent of economic policy reform.

We provide a detailed discussion of the estimated absorptive capacity evolution in its relation to economic poliy in three countries which experienced very different trajectories over the past six decades (Ireland, Japan, and Sweden). In order to generalise the insights from this exercise we further provide indications of the relationship between our absorptive capacity estimates and indicators for financial development, human capital and competition policy in our wider sample of countries.

The remainder of this study is structured as follows: the following Section sets out the empirical model and demonstrates how this can be interpreted as an unobserved component model in state space form. It then sketches the empirical implementation using Bayesian simulation-based methods. The data along with the empirical results are discussed in Section 3. In Section 4 we link our findings to economic policy via case studies and various descriptive analyses. Section 5 concludes.

2 Empirical Specification and Implementation

We present a factor-augmented Cobb-Douglas production function with time-varying absorptive capacity and suggest a CCE approach to identify unobserved global technology. Transformed into a state space model our empirical model can further be employed to explicitly test for time-variation in the levels and growth effects of absorptive capacity. We adopt a Markov Chain Monte Carlo (MCMC) approach to estimate the model.

2.1 Empirical Model

We model output in country $i = 1, \dots, N$ at time $t = 1, \dots, T$ using a Cobb-Douglas production function with constant returns to scale

$$Y_{it} = \Lambda_{it} K_{it}^{\beta_i} L_{it}^{1-\beta_i} e^{\epsilon_{it}}, \qquad \text{with} \quad 0 < \beta_i < 1 \quad \forall i, \tag{1}$$

where Y_{it} is real GDP, K_{it} is real private capital stock, L_{it} is total hours worked and ϵ_{it} is a zero-mean stationary error term uncorrelated across countries. To allow for a heterogeneous production function, β_i is a random coefficient with fixed mean and finite variance. Unobserved TFP, Λ_{it}, is defined broadly as the intangible technology and knowledge stock but also to incorporate the effects of human capital and public infrastructure among other factors.

Common factor structure

Building on an established strand of the literature that considers a country's TFP to be the successful assimilation of global technology (Parente and Prescott, 1994, 2002; Alfaro et al., 2008), we parameterize Λ_{it} using a common factor framework

$$\Lambda_{it} = \mathcal{A}_{it} F_t^{\vartheta_{it}}, \tag{2}$$

where F_t is a common factor that we interpret as representing the worldwide available technology and knowledge stock while $\mathcal{A}_{it}$ and ϑ_{it} capture country-specific endowments, institutions, investments and policies that determine how much of F_t is successfully appropriated (henceforth 'absorptive capacity'). Substituting equation (2) in (1), dividing by hours worked L_{it} and taking logarithms yields

$$y_{it} = (a_{it} + \vartheta_{it} f_t) + \beta_i k_{it} + \epsilon_{it}, \tag{3}$$

where $a_{it} = ln\,(\mathcal{A}_{it})$, $y_{it} = ln\,(Y_{it}/L_{it})$, $k_{it} = ln\,(K_{it}/L_{it})$, $f_t = ln\,(F_t)$.

Changes in absorptive capacity: level versus growth shifts

The empirical specification in equation (3) is closely related to that of Eberhardt et al. (2013), who use a common factor framework with time-invariant parameters, and Everaert et al. (2014), who allow absorptive capacity to vary as a function of fiscal policy variables. Our main contribution is to allow for a flexible evolution in a_{it} and ϑ_{it}, and hence in absorptive capacity, *over time*. Although our setup is notable for the absence of any explicit mechanism for knowledge creation, this empirical specification for Λ_{it} is a generalization of the one presented in the growth model of Klenow and Rodríguez-Clare (2005) where policies and other efforts to improve absorptive capacity only have a level effect on TFP. The main result of their model is that in the long run all countries share a common growth rate equal to the growth rate of global TFP – a result they empirically motivate by demonstrating that countries with high investment rates typically have higher levels of wealth rather than higher growth rates. In our model this equates to setting $\vartheta_{it} = 1$. In order to allow for an endogenous type of growth where country-specific characteristics or policies can have permanent growth effects we allow for the possibility that $\vartheta_{it} \neq 1$ and that it varies across countries and time. Hence, the advantage of the exponential common factor structure for Λ_{it} in equation (2) is that it allows us to distinguish between advances in absorptive capacity that lead to *level* versus *growth* shifts in a country's TFP. To illustrate this, it is convenient to look at a Taylor expansion of Λ_{it}

$$\Lambda_{it} = e^{a_{it}+\vartheta_{it}f_t} = (1 + a_{it}) + (1 + a_{it})\,\vartheta_{it}f_t + \ldots, \tag{4}$$

together with the growth rate of Λ_{it}

$$\Delta \ln \Lambda_{it} = \Delta a_{it} + \Delta\vartheta_{it}f_{t-1} + \vartheta_{it}\Delta f_t. \tag{5}$$

In the absence of changes in a_{it} and ϑ_{it}, the growth of a country's TFP is a fixed proportion $\vartheta_{it} = \vartheta_i$ of the growth rate of global TFP Δf_t: global knowledge has permanent growth effects but their magnitude is 'predetermined' and not subject to policy intervention. Equation (4) shows that an increase in a_{it} implies that a country is able to assimilate more of the global technology f_t, while from equation (5) it is clear that this leaves the future growth rate of the economy unaffected. Hence, advances in absorptive capacity that lead to a level shift in a country's TFP will be captured by changes in a_{it}. A shock to ϑ_{it} induces a similar levels shift but, as apparent from the term $\vartheta_{it}\Delta f_t$ in equation (5), this also implies that the economy will now grow at a *permanently* higher rate.

2.2 CCE approach to identify unobserved worldwide technology

One possible way to identify unobserved worldwide technology f_t in equation (3) is to write down a data generating process (e.g. a random walk with time-varying drift), cast the model in state space form and filter f_t using the Kalman filter. An important challenge in this approach is that only the product $\vartheta_{it}f_t$ is identified but not the constituent components: multiplying the loadings ϑ_{it} by a rescaling constant c while dividing the common factor f_t by the same c would leave the product unchanged. A standard normalization is therefore to constrain the scale of the factor f_t. However, while being effective in a model with fixed loadings, time variation brings about a new identification issue as the rescaling term can now be a time-varying sequence c_t rather than a constant c. A further identification problem may arise when separating the idiosyncratic components a_{it} and ϑ_{it}. Although these are assumed to be uncorrelated across countries, this is not explicitly imposed by the Kalman filter that will be used to estimate them, such that there is some scope for a_{it} to pick up common technology trends that should in fact be captured by $\vartheta_{it} f_t$.

In this paper, we therefore follow a different route to identify the common factor f_t and the time-varying absorptive capacity parameters a_{it} and ϑ_{it}. Inspired by the CCE approach of Pesaran (2006), taking cross-sectional averages of the model in equation (3) yields

$$\overline{y}_t = \overline{a}_t + \overline{\vartheta}_t f_t + \overline{\beta}\,\overline{k}_t + \overline{\epsilon}_t, \tag{6}$$

where $\overline{y}_t = \frac{1}{N}\sum_{i=1}^{N} y_{it}$ and similarly for the other variables. Solving for f_t

$$f_t = \frac{1}{\overline{\vartheta}_t}\left(\overline{y}_t - \overline{a}_t - \overline{\beta}\,\overline{k}_t - \overline{\epsilon}_t\right), \tag{7}$$

and substituting this solution back into equation (3) yields

$$y_{it} = \left(a_{it} - \frac{\vartheta_{it}}{\overline{\vartheta}_t}\overline{a}_t\right) + \frac{\vartheta_{it}}{\overline{\vartheta}_t}\left(\overline{y}_t - \overline{\beta}\,\overline{k}_t\right) + \beta_i k_{it} + \left(\epsilon_{it} - \frac{\vartheta_{it}}{\overline{\vartheta}_t}\overline{\epsilon}_t\right), \tag{8}$$

$$= \alpha_{it} + \theta_{it}\widehat{f}_t + \beta_i k_{it} + \varepsilon_{it}, \tag{9}$$

where $\alpha_{it} = a_{it} - \frac{\vartheta_{it}}{\overline{\vartheta}_t}\overline{a}_t$, $\theta_{it} = \frac{\vartheta_{it}}{\overline{\vartheta}_t}$, $\widehat{f}_t = \overline{y}_t - \overline{\beta}\,\overline{k}_t$ and $\varepsilon_{it} = \epsilon_{it} - \frac{\vartheta_{it}}{\overline{\vartheta}_t}\overline{\epsilon}_t$. Given the assumption that ϵ_{it} is a zero-mean white noise term uncorrelated across cross-sections, we have that $\overline{\epsilon}_t \xrightarrow{p} 0$ as $N \to \infty$ such that, conditional on the capital elasticity parameters β_i that determine $\overline{\beta}$, equation (7) implies that $\widehat{f}_t$ can be used as an observable proxy for the rescaled and recentred factor $\overline{\vartheta}_t f_t - \overline{a}_t$.

It is easily verified that the normalizations imposed when going from equation (1) to (9) solve the identification issues outlined at the beginning of this section. First, the scale of $\widehat{f}_t$ is determined by that of the cross-sectional averages $\overline{y}_t$ and $\overline{k}_t$. Second, the factor loadings θ_{it} are normalised to be one on average across countries in every period, i.e. $\frac{1}{N}\sum_{i=1}^{N}\theta_{it} = \frac{1}{N}\sum_{i=1}^{N}\vartheta_{it}/\overline{\vartheta}_t = 1\ \forall t$, such that they can no longer be multiplied by a time-varying sequence c_t. Third, the cross-sectional average of a_{it} is normalised to zero in every period, i.e. $\frac{1}{N}\sum_{i=1}^{N}\alpha_{it} = \frac{1}{N}\sum_{i=1}^{N}(a_{it} - \frac{\vartheta_{it}}{\overline{\vartheta}_t}\overline{a}_t) = 0\ \forall t$, such that it cannot pick up common technology trends. Note that our normalizations imply that $\widehat{f}_t$ should not be interpreted as the world TFP frontier but rather as an index of average world technology, with the combination of α_{it} and θ_{it} indicating whether a country operates below or above this average level.

2.3 Modelling and testing for time-varying absorptive capacity

At the heart of our paper are the time-varying parameters α_{it} and θ_{it} measuring a country's efficiency to incorporate the world technology into its own production techniques. This time variation implies that equation (9) cannot be estimated using the standard CCE approach of Pesaran (2006). As an alternative, we set up a state space model. We further use a Bayesian model specification search procedure to analyse whether our generalization to a time-varying parameters setting is empirically relevant. If the restrictions $\alpha_{it} = \alpha_i$ and $\theta_{it} = \theta_i$ are valid, our model simplifies to a standard common factor error structure that can be estimated using the conventional CCE approach.

State space model

We complete the model by assuming that the absorptive capacity parameters α_{it} and θ_{it} evolve according to random walk processes

$$\alpha_{it} = \alpha_{it-1} + \psi_{it}^{\alpha}, \qquad \psi_{it}^{\alpha} \sim N(0, \varsigma_{\alpha}), \tag{10}$$

$$\theta_{it} = \theta_{it-1} + \psi_{it}^{\theta}, \qquad \psi_{it}^{\theta} \sim N(0, \varsigma_{\theta}). \tag{11}$$

The random walk assumption allows for a very flexible evolution of the parameters over time. The model can then be cast in its state space representation with (9) being the 'observation equation', where for the noise term ε_{it} we assume $\varepsilon_{it} \sim N\left(0, \sigma_{\varepsilon}^2\right)$, and (10)-(11) the 'state equations' such that the random walk components α_{it} and θ_{it} can be estimated using the Kalman filter.

Bayesian stochastic model specification search

Determining whether the proposed time-variation in the parameters α_{it} and θ_{it} is relevant implies testing if the innovation variances ς_{α} and ς_{θ} in equations (10)-(11) are zero or not. From a classical point of view this is cumbersome as the null hypothesis of a zero variance lies on the boundary of the parameter space. We therefore use the stochastic model specification search of Frühwirth-Schnatter and Wagner (2010), generalizing standard Bayesian variable selection to state space models. This involves reparametrising the state equations (10)-(11) to:

$$\alpha_{it} = \alpha_{i0} + \sqrt{\varsigma_{\alpha}}\widetilde{\alpha}_{it}, \qquad \text{with } \widetilde{\alpha}_{it} = \widetilde{\alpha}_{i,t-1} + \widetilde{\psi}_{it}^{\alpha}, \qquad \widetilde{\alpha}_{i0} = 0, \qquad \widetilde{\psi}_{it}^{\alpha} \sim N\left(0,1\right), \tag{12}$$

$$\theta_{it} = \theta_{i0} + \sqrt{\varsigma_{\theta}}\widetilde{\theta}_{it}, \qquad \text{with } \widetilde{\theta}_{it} = \widetilde{\theta}_{i,t-1} + \widetilde{\psi}_{it}^{\theta}, \qquad \widetilde{\theta}_{i0} = 0, \qquad \widetilde{\psi}_{it}^{\theta} \sim N\left(0,1\right), \tag{13}$$

which splits α_{it} and θ_{it} into initial values α_{i0} and θ_{i0} and the (possibly) time-varying parts $\sqrt{\varsigma_{\alpha}}\widetilde{\alpha}_{it}$ and $\sqrt{\varsigma_{\theta}}\widetilde{\theta}_{it}$.

This 'non-centered' parametrization has a number of interesting features. First, the signs of both $\sqrt{\varsigma_{\alpha}}$ and $\widetilde{\alpha}_{it}$ can be changed without changing their product, and similarly for $\sqrt{\varsigma_{\theta}}$ and $\widetilde{\theta}_{it}$. This lack of identification offers a first piece of information about whether time-variation is relevant or not: for truly time-*varying* parameters, the innovation variance ς will be positive resulting in a posterior distribution of $\sqrt{\varsigma}$ that is bimodal with modes $\pm\sqrt{\varsigma}$. For time-*invariant* parameters, ς is zero such that $\sqrt{\varsigma}$ becomes unimodal at zero.

Second, the non-centered parametrization is very useful for model selection as it represents α_{it} and θ_{it} as a superposition of the initial values α_{i0} and θ_{i0} and the time-varying components $\widetilde{\alpha}_{it}$ and $\widetilde{\theta}_{it}$. As a result, in contrast to the centered parametrization in equations (10)-(11), $\widetilde{\alpha}_{it}$ and $\widetilde{\theta}_{it}$ do not degenerate to a static component when the innovation variances are zero. In fact, when for instance $\varsigma_\alpha \approx 0$, then $\sqrt{\varsigma_\alpha} \approx 0$ and $\widetilde{\alpha}_{it}$ will drop from the model. As suggested by Frühwirth-Schnatter and Wagner (2010), this allows us to cast the test of whether the variances ς_α and ς_θ are zero or not into a more regular variable selection problem. To this end we introduce two binary indicator variables δ_α and δ_θ, which are equal to one if the corresponding parameter varies over time and zero otherwise. The resulting parsimonious non-centered specification is then given by

$$y_{it} = \left[(\alpha_{i0} + \delta_\alpha \sqrt{\varsigma_\alpha}\widetilde{\alpha}_{it}) + \left(\theta_{i0} + \delta_\theta \sqrt{\varsigma_\theta}\widetilde{\theta}_{it}\right) \widehat{f}_t\right] + \beta_i k_{it} + \varepsilon_{it}. \tag{14}$$

When $\delta_\alpha = 1$, α_{i0} is the initial value of α_{it} and $\sqrt{\varsigma_\alpha}$ is an unconstrained parameter that is estimated from the data. Alternatively, when $\delta_\alpha = 0$ the time-varying part $\widetilde{\alpha}_{it}$ drops out and α_{i0} represents the time-invariant parameter.

A third important advantage of the non-centered parametrization is that it allows us to replace the standard Inverse Gamma prior on the variance parameters ς_α and ς_θ by a Gaussian prior centered at zero on $\sqrt{\varsigma_\alpha}$ and $\sqrt{\varsigma_\theta}$. Centering the prior distribution at zero is possible as for both $\varsigma = 0$ and $\varsigma > 0$, $\sqrt{\varsigma}$ is symmetric around zero, with the main difference being that in the latter case the posterior distribution is bimodal.[4]

2.4 Mean Group versus Pooled estimators

The model outlined above allows the capital elasticity coefficient β_i to be heterogeneous over cross-sections. If the parameters of interest are the cross-sectional means rather than the heterogeneous values, we consider two alternative ways to pool the estimates. In line with Pesaran (2006), the first approach is to calculate simple cross-sectional averages of the individual coefficient, while the second is to impose the homogeneity restriction that $\beta_i = \beta$. We will refer to these as Mean Group (MG) and Pooled estimators, respectively. Note that the parameters α_{it} and θ_{it}, and their constituent components $(\alpha_{i0}, \widetilde{\alpha}_{it})$ and $(\theta_{i0}, \widetilde{\theta}_{it})$, are always fully heterogeneous.

2.5 MCMC algorithm

The state space representation in equations (12)-(14) is a non-linear model for which the standard approach of using the Kalman filter to obtain the time-varying components and Maximum Likelihood to estimate the unknown parameters is inappropriate. We therefore use an MCMC approach to jointly sample the binary indicators $\boldsymbol{\delta} = \{\delta_\alpha, \delta_\theta\}$, the unrestricted elements of the parameter vector $\boldsymbol{\phi} = \{\alpha_{i0}, \theta_{i0}, \sqrt{\varsigma}, \beta_i, \sigma^2_\varepsilon\}_{i=1}^N$ and the latent state processes $\boldsymbol{s} = \{\{\widetilde{\alpha}_{it}, \widetilde{\theta}_{it}\}_{t=1}^T\}_{i=1}^N$ from the posterior distribution $g(\boldsymbol{\delta}, \boldsymbol{\phi}, \boldsymbol{s}|\boldsymbol{x})$ conditional on the data $\boldsymbol{x} = \{\{y_{it}, k_{it}\}_{t=1}^T\}_{i=1}^N$. This conveniently splits the

[4] Frühwirth-Schnatter and Wagner (2010) show that compared to using an Inverse Gamma prior for ς, the posterior density of $\sqrt{\varsigma}$ is much less sensitive to the hyperparameters of the Gaussian distribution and, importantly, is not pushed away from zero when $\varsigma = 0$.

non-linear estimation problem into a sequence of blocks which are linear conditional on the other blocks. Given a set of starting values, sampling from the various blocks is iterated K times and, after a sufficiently long burn-in period B, the sequence of draws $(B+1,...,K)$ approximates a sample from $g(\boldsymbol{\delta},\boldsymbol{\phi},\boldsymbol{s}|\boldsymbol{x})$. The results reported below will be based on $K=45,000$ with the first $B=5,000$ discarded as burn-in. Following Frühwirth-Schnatter and Wagner (2010), we fix the binary indicators in $\boldsymbol{\delta}$ to be one during the first $2,500$ iterations of the burn-in period to obtain sensible starting values for the unrestricted model before variable selection actually starts. A detailed description of the different blocks together with an interweaving approach to boost the mixing efficiency of the MCMC algorithm is presented in Appendix A.

3 Data, Coefficient Priors and Results

3.1 Data

We estimate our empirical specification for a panel of 31 predominantly high-income countries using annual data over the period 1953-2014 taken from the Penn World Table (PWT) version 9 (Feenstra et al., 2015) – our sample is made up of 26 current OECD member countries (comprising all current members with the exception of Israel, Turkey and the seven former transition economies), with the addition of Argentina, Brazil, Colombia, Cyprus and Taiwan.[5] At the start of our sample these 31 countries account for over 80% of world GDP (measured in PPP terms at constant 2011 national prices), declining to just under half in 2014. Table 1 outlines details on the data construction. Real GDP and the real capital stock are in constant 2011 national prices transformed into 2011 US$. The capital stock defined by PWT version 9 includes residential structures. Total hours worked is calculated by multiplying the number of persons engaged times the average annual hours worked per person.

Table 1: Construction of data and data sources

Name	Notation	Construction	Code
Real GDP (in million US$, 2005 values)	Y_{it}	PWT data	rgdpna
Real Capital stock (in million US$, 2005 values)	K_{it}	PWT data	rkna
Number of persons engaged	N_{it}	PWT data	emp
Average annual hours worked by persons engaged	H_{it}	PWT data	avh
Total annual hours worked in the economy	L_{it}	$N_{it} \times H_{it}$	
Log output per hour worked	y_{it}	$log\,(Y_{it}\,/L_{it})$	
Log capital per hour worked	k_{it}	$log\,(K_{it}\,/L_{it})$	

[5]The number of countries is purely driven by data availability where we include all countries for which we have 62 years of data resulting in a balanced panel.

3.2 Prior choices

For the variance σ_ε^2 of the errors terms ε_{it} we use an uninformative Inverse Gamma prior distribution $IG(c_0, C_0)$, with the shape c_0 and scale C_0 parameters both set to 0.001. For all other parameters we use a Gaussian prior distribution $N\left(a_0, A_0\sigma_\varepsilon^2\right)$ defined by setting a prior belief a_0 with prior variance $A_0\sigma_\varepsilon^2$. Throughout the estimation procedure we fix A_0 to 100^2 to ensure that posterior results are driven by the information contained in the data. Note that the non-centered parametrization in equation (14) allows us to make use of a Normal prior for the standard deviations $\sqrt{\varsigma_\alpha}$ and $\sqrt{\varsigma_\theta}$ since estimating them boils down to a standard linear regression. When sampling the indicators $\boldsymbol{\delta}$ we assign 50%, i.e. $p_0 = 0.5$, prior probability that the indicators take on the value 1. The Pooled and MG estimators are based on the same priors. The prior beliefs are chosen as follows:

- **Output elasticity with respect to capital** β_i. We follow the common assumption in the literature that this elasticity should lie in the neighbourhood of 0.33 (Gollin, 2002).
- **Initial values and non-centered components** α_{i0}**,** θ_{i0}**,** $\sqrt{\varsigma_i^\alpha}$**,** $\sqrt{\varsigma_i^\theta}$. For the initial values of the time-varying processes, α_{i0} and θ_{i0}, our prior beliefs are 0 and 1, respectively. This is a natural outcome of the way the CCE approach normalises α_{it} and θ_{it}, i.e. $\frac{1}{N}\sum_{i=1}^N \alpha_{it} = 0$ and $\frac{1}{N}\sum_{i=1}^N \theta_{it} = 1$ for all t and hence also for the initial values α_{i0} and θ_{i0}. Our prior belief for $\sqrt{\varsigma_i^\alpha}$ and $\sqrt{\varsigma_i^\theta}$ is 0. We center this distribution around zero such that our belief is in accordance with the null hypothesis of our test for whether α_{it} and θ_{it} vary or are fixed over time.

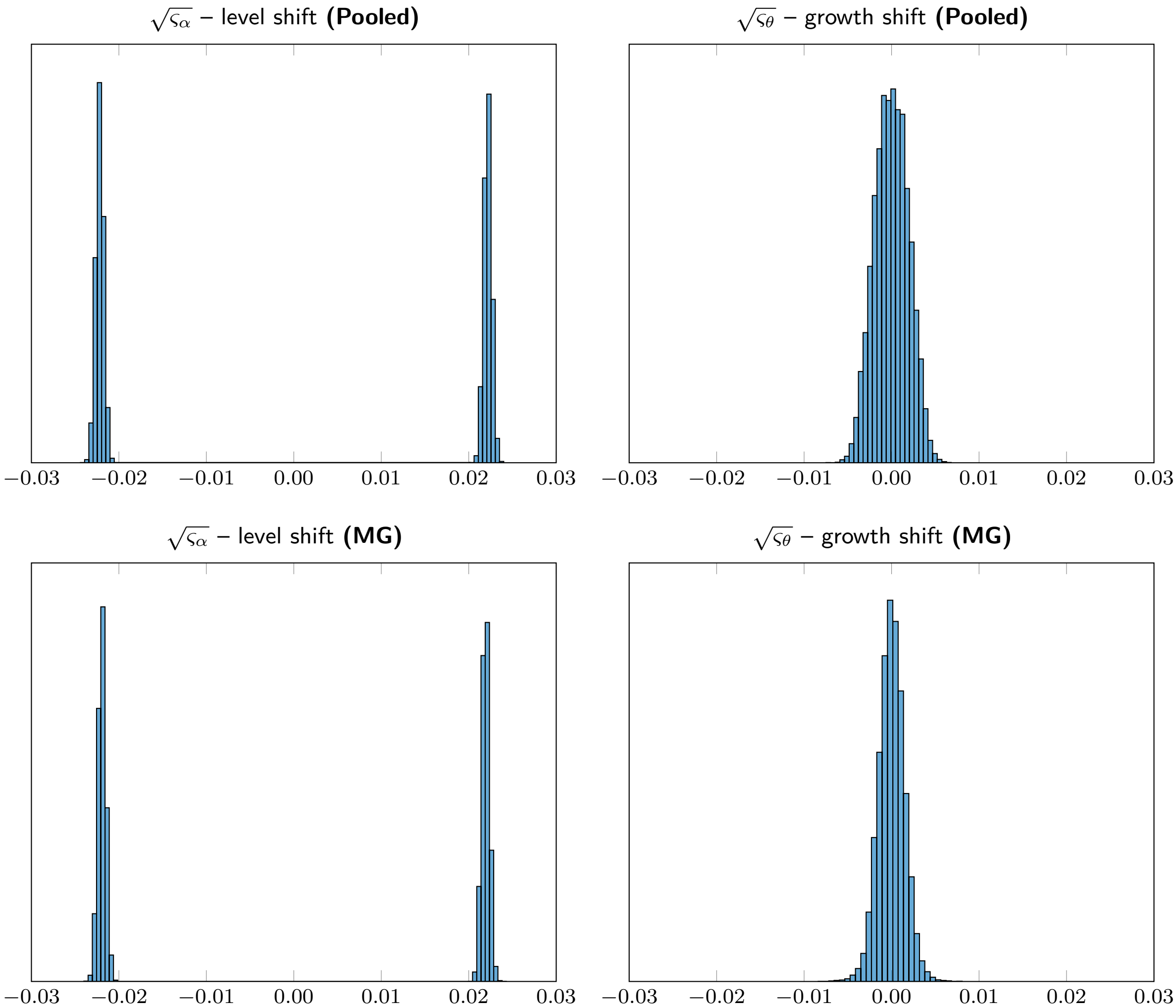
3.3 Empirical results

Time variation in the absorptive capacity parameters

We start by discussing the results of the stochastic model specification search used to analyse whether time variation in the absorptive capacity parameters α_{it} and θ_{it} is a relevant aspect of the model. This enables us to discriminate between the four possible models nested in our set-up, i.e. a model where changes in absorptive capacity lead to either growth or level shifts in TFP, a combination of the two or a model without any changes in absorptive capacity.

As a first step, we fix the binary indicators in $\boldsymbol{\delta}$ to 1 to obtain posterior distributions for the unrestricted model where both α_{it} and θ_{it} are allowed to vary over time. When time variation is relevant, this should show up as bimodality in the posterior distribution of the corresponding innovation standard deviation $\sqrt{\varsigma}$. A unimodal distribution centered at zero is expected for time-invariant parameters. Figure 1 plots the posterior distributions of $\sqrt{\varsigma_\alpha}$ and $\sqrt{\varsigma_\theta}$ for both the Pooled and the MG estimator. The results are decisive in that the posterior distribution of $\sqrt{\varsigma_\alpha}$ shows clear bimodality while that of $\sqrt{\varsigma_\theta}$ is perfectly unimodal. This already offers a strong indication that the information in the data answers to a model with level but no growth shifts in TFP.

Figure 1: Posterior distributions of $\sqrt{\varsigma_\alpha}$ and $\sqrt{\varsigma_\theta}$



As a more formal test for time variation, we sample the stochastic binary indicators in $\boldsymbol{\delta}$ together with the other parameters in the model. Table 2 reports the posterior probabilities for the binary indicators being one, calculated as the fraction of draws in which the stochastic model specification search prefers a model which allows for time variation in the corresponding parameter. We also report posterior probabilities for each of the four models that can be formed as combinations of the two binary indicators. It is clear that time variation is important as the model with $\delta_\theta = \delta_\alpha = 0$ has zero probability of being selected. The finding that the pooled estimation procedure assigns a 91% probability to the model $(\delta_\alpha, \delta_\theta) = (1, 0)$ further supports our previous conclusion that in particular α_{it} exhibits relevant time variation while θ_{it} is most likely constant over time. A similar conclusion can be drawn when considering the MG estimator. Taken together this suggests that in our sample we find evidence *against* a model where the long-run growth rate of TFP can be altered using policy interventions.

Table 2: Posterior inclusion probabilities for the binary indicators $\boldsymbol{\delta}$ and their combinations

	Models (δ_α,δ_θ)				Indicators	
	(0,0)	(1,1)	(1,0)	(0,1)	δ_α	δ_θ
Pooled	0.00	0.09	0.91	0.00	1.00	0.09
MG	0.00	0.10	0.90	0.00	1.00	0.10

Klenow and Rodríguez-Clare (2005)

Based on the stochastic model specification search, we can conclude that there is relevant time variation in α_{it} but not in θ_{it}. The latter still allows for $\theta_i \neq 1$, whereas an intrinsic property of the model put forward by Klenow and Rodríguez-Clare (2005) is that $\theta_i = 1$, such that in the long run all countries grow at the same pace. Table 3 reports posterior results for θ_i obtained from estimating a parsimonious specification where we set $\delta^\alpha = 1$ and $\delta^\theta = 0$. For most but not all countries 1 is included in the 90% highest density interval (HDI). In order to test this in a more rigorous way, we can again use the stochastic variable selection approach. To this end we (i) split θ_i into 1 and its deviation $(\theta_i - 1)$, and (ii) add a binary indicator $\gamma_{i\theta}$ that equals one when the corresponding variable $(\theta_i - 1)\widehat{f}_t$ should be included in the model and zero otherwise. This results in the following specification

$$y_{it} - \widehat{f}_t = \alpha_{it} + \gamma_{i\theta}(\theta_i - 1)\widehat{f}_t + \beta_i k_{it} + \varepsilon_{it}, \tag{15}$$

where for $\gamma_{i\theta} = 1$ the deviation of θ_i from one is estimated from the data while $\gamma_{i\theta} = 0$ implies that θ_i is set to one. Table 3 reports the posterior probability that $(\theta_i - 1)\widehat{f}_t$ should enter the model calculated as the frequency $\gamma_{i\theta}$ takes on the value of one over the MCMC iterations. For most countries the results are in line with the model of Klenow and Rodríguez-Clare (2005) as deviations of θ_i from one are not found to be a relevant aspect of the model. However, for a number of countries the restriction that $\theta_i = 1$ is not supported by the data. This is most prominently the case for Australia, Cyprus and Taiwan, for which the posterior model inclusion probability of $(\theta_i - 1)\widehat{f}_t$ clearly exceeds 50% for both the Pooled and the MG estimator, and to a lesser extent also for Brazil, Canada, Greece, Luxembourg, Portugal and South Korea, where the posterior model inclusion probability of $(\theta_i - 1)\widehat{f}_t$ clearly exceeds 50% for either the Pooled or the MG estimator. This suggests that over the period 1953-2014 a number of countries in our dataset did have TFP growth that differed from the global evolution. However, we need to point out that the aforementioned countries typically are those that have caught-up to (Brazil, Cyprus, Greece, Portugal, Taiwan) or have been caught-up by (Canada) the global TFP evolution. As far as our sample covers a prolonged period of catching-up, this effect may result in $\theta_i \neq 1$ instead of showing up as time-variation in α_{it}. A longer sample may be needed to rule out this possibility.

Table 3: Posterior results for θ_i and $\gamma_{i\theta}$

	θ_i				$\gamma_{i\theta}$	
	Pooled		MG		Pooled	MG
Argentina	0.77	(0.18)	0.95	(0.19)	0.35	0.24
Australia	0.61	(0.18)	0.47	(0.20)	0.74	0.85
Austria	1.18	(0.18)	1.03	(0.23)	0.29	0.17
Belgium	1.02	(0.18)	0.89	(0.22)	0.22	0.18
Brazil	1.27	(0.18)	1.29	(0.19)	0.46	0.63
Canada	0.60	(0.18)	0.65	(0.20)	0.73	0.48
Chili	0.73	(0.18)	0.97	(0.21)	0.40	0.38
Colombia	0.76	(0.18)	0.76	(0.20)	0.39	0.45
Cyprus	1.66	(0.18)	1.91	(0.21)	0.99	1.00
Denmark	0.72	(0.18)	0.72	(0.22)	0.48	0.28
Finland	1.20	(0.18)	1.38	(0.22)	0.35	0.44
France	1.06	(0.18)	0.87	(0.23)	0.22	0.19
Germany	1.11	(0.18)	0.94	(0.24)	0.26	0.23
Greece	1.42	(0.18)	1.39	(0.22)	0.79	0.56
Iceland	0.79	(0.18)	0.74	(0.22)	0.35	0.28
Ireland	1.21	(0.18)	1.15	(0.20)	0.37	0.34
Italy	1.14	(0.18)	1.05	(0.23)	0.29	0.21
Japan	0.75	(0.19)	1.36	(0.22)	0.44	0.54
Luxembourg	1.29	(0.18)	1.56	(0.21)	0.52	0.83
Mexico	0.97	(0.18)	0.95	(0.20)	0.23	0.37
Netherlands	0.93	(0.18)	0.87	(0.22)	0.23	0.24
New Zealand	0.91	(0.18)	0.82	(0.21)	0.24	0.28
Norway	0.92	(0.18)	0.70	(0.22)	0.25	0.29
Portugal	1.35	(0.18)	1.31	(0.23)	0.66	0.26
South Korea	0.87	(0.18)	0.70	(0.20)	0.30	0.78
Spain	1.16	(0.18)	1.06	(0.21)	0.32	0.18
Sweden	0.78	(0.18)	0.83	(0.22)	0.38	0.28
Switzerland	0.72	(0.18)	0.92	(0.22)	0.45	0.19
Taiwan	1.61	(0.18)	1.53	(0.21)	0.98	0.91
UK	0.77	(0.18)	0.74	(0.21)	0.36	0.27
USA	0.72	(0.18)	0.73	(0.20)	0.46	0.40

[a] Standard deviations of the posterior distributions are reported in parentheses.
[b] Posterior results for θ_i are obtained from a parsimonious specification where we restrict δ^α and δ^θ based on the outcome of the stochastic model specification search, i.e. $\delta^\alpha = 1$ and $\delta^\theta = 0$. The last two columns effectively report the probability that $\theta_i \neq 1$.
[c] Based on MCMC with $K = 45,000$ iterations where the first $B = 5,000$ are discarded as burn-in.

Production function estimates

Table 4 reports posterior results for the parameters in the production function. Apart from the unrestricted model in [1], where we impose $\delta_\alpha = \delta_\theta = 1$, in [2] we also present results for a parsimonious specification where based on the outcome of the stochastic model specification search we set $\delta^\alpha = 1$ and $\delta^\theta = 0$, such that $\theta_{it} = \theta_i$. We further estimate a hybrid parsimonious specification in [3] where θ_i is restricted to one for those countries where the posterior probability that $\gamma_{i\theta} = 1$ is below 0.5. Finally, in [4] we estimate a restricted model where $\theta_i = 1$ for all countries. The estimated output elasticity with respect to capital β is found to be slightly above 0.5 for both the Pooled and the MG estimator across all four specifications (country-specific estimates for β_i are reported in Appendix B). As expected, there is a lot of uncertainty around the individual β_i's, which also spills over to the other parameters in the model. Since the assumption of common technology for advanced economies is quite uncontroversial we restrict our discussion below to the results for the Pooled estimator.

Table 4: Production function estimates for different estimators

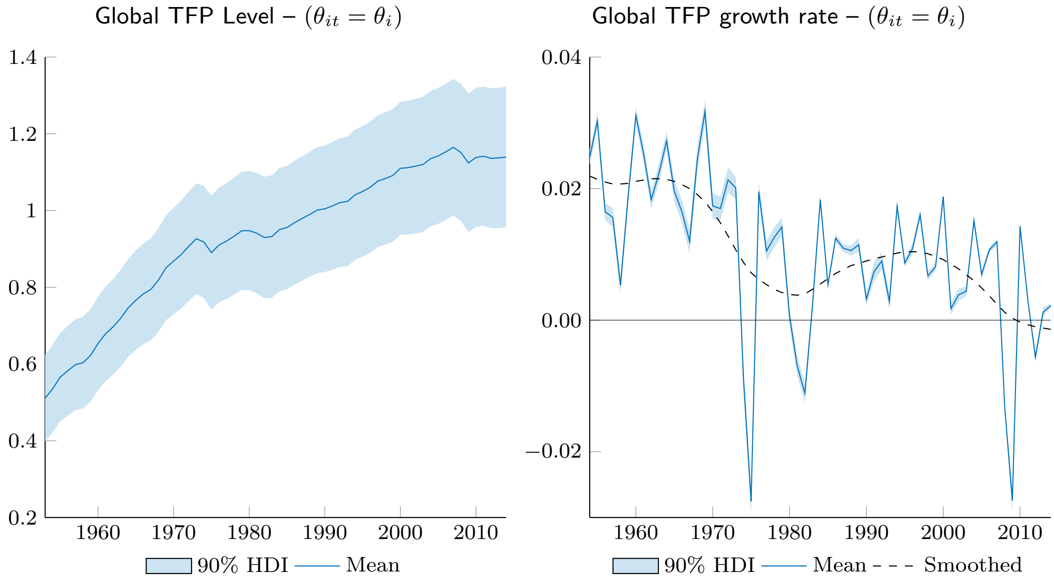
	[1] Unrestricted		[2] Parsimonious $(\theta_{it} = \theta_i)$		[3] Parsimonious – hybrid		[4] Parsimonious $(\theta_i = 1)$	
	Pooled	MG	Pooled	MG	Pooled	MG	Pooled	MG
$\sqrt{\varsigma_\alpha}$	0.0222	0.0219	0.0222	0.0219	0.0226	0.0219	0.0230	0.0222
	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
$\sqrt{\varsigma_\theta}$	0.0016	0.0012						
	(0.0011)	(0.0009)						
β	0.51	0.56	0.50	0.56	0.51	0.50	0.52	0.54
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
σ_ε	0.0072	0.0067	0.0072	0.0067	0.0063	0.0070	0.0056	0.0070

[a] Standard deviations of the posterior distributions are reported in parentheses.
[b] In the hybrid model we set $\theta_i = 1$ only for the countries where the posterior probability $\gamma_{i\theta} = 1$ is smaller than 0.5.

Estimated global TFP

Figure 2 plots posterior results for the global TFP index $\widehat{f_t}$ and its growth rate using the parsimonious specification $(\theta_{it} = \theta_i)$. The results when restricting $\theta_i = 1$ are close to identical. In line with the productivity growth patterns documented by Blanchard (2004), Madsen (2008) and van Ark et al. (2008), our results show that the post-war period can be split into three episodes. First, the 1950s and 1960s are a period of high and stable global TFP growth in excess of 2% per annum. Second, the early 1970s show a steep decline in the global productivity growth rate, heralding an era of lower growth. A decline in TFP growth over time can be squared with the observation of declining R&D expenditure growth in the major economies in our OECD country sample over the 1980s and 1990s (see Appendix C). Third, a slight improvement during the 1990s and early 2000s whereafter global TFP growth nosedives again during the Global Financial Crisis (GFC) in 2007/8 and seems to stabilize around 0% afterwards. This raises the question whether the GFC signals a new era of stagnant global TFP – with the data restrictions on time since the GFC we are unable to address this in the present study.

Figure 2: Posterior global TFP level $\widehat{f}_t$ and its growth rate



Absorptive capacity evolution

Figure 3 presents posterior results for the time-varying absorptive capacity parameter α_{it} for the parsimonious specification setting $\theta_{it} = \theta_i$ (in blue) and the restricted model where $\theta_i = 1$ (in red). The evolution in α_{it} is very similar for both settings, but the precision of the estimates is much lower when estimating θ_i while the levels of α_{it} differ as well, especially for those countries where $\theta_i = 1$ is not supported by the data. This shows that given the information available in our sample, it is difficult to assign deviations of country-specific TFP from the global level to $\alpha_{it} \neq 0$ or to $\theta_i \neq 1$. The source of this difficulty can be illustrated via the normalised and restricted $(\theta_{it} = \theta_i)$ version of equation (4)

$$\Lambda_{it} = e^{\alpha_{it} + \theta_i \widehat{f}_t} = (1 + \alpha_{it}) + (1 + \alpha_{it})\, \theta_i \widehat{f}_t + \ldots$$

A reduction in θ_i can be compensated by an increase in the average level of α_{it} leaving the first-order impact of $\widehat{f}_t$ on country-specific TFP Λ_{it} unchanged. Hence, identification stems from the level term $(1 + \alpha_{it})$ and from the higher order effects of $\widehat{f}_t$ on Λ_{it}. Restricting $\theta_i = 1$ results in much less uncertainty around the level of α_{it}. A normalised and restricted $(\theta_{it} = \theta_i)$ version of equation (5)

$$\Delta \ln \Lambda_{it} = \Delta a_{it} + \theta_i \Delta f_t,$$

further shows that it is much easier to separately identify $\Delta\alpha_{it}$ and θ_i.

Looking at the evolution in the absorptive capacity parameter α_{it} for either $\theta_{it} = \theta_i$ or $\theta_{it} = 1$, there is substantial variation over time in many countries. A first group, including Cyprus, Finland, Ireland, South Korea and Taiwan, show an increase in their ability to assimilate foreign knowledge. These countries are clearly catching up with the rest since they started off well below average absorptive capacity in 1953. The opposite evolution can be observed for a second group, consisting of Japan and Switzerland, and to a lesser extent Argentina. These countries start with clearly positive relative absorptive capacity parameters α_{it}, but exhibit a (secular) decline over the sample period. Other countries show either a modest increase or decrease, with Australia, Austria, Denmark, France and the Netherlands showing little or no structural movement in α_{it}. The seemingly 'static' nature of the latter group of countries is however somewhat misleading, as would be the same verdict for the global technology leader, the United States, which saw a mild increase in α_{it}: recall that the absorptive capacity evolution charted here is a relative index, such that these countries can be highlighted as having kept up a very strong absorptive capacity performance consistently on par with (Austria, Belgium, France, among others) or even outpacing the global developments (Australia, Canada, Denmark, Sweden, Norway and the U.S.) over this time period.

4 Linking absorptive capacity evolution and economic policy

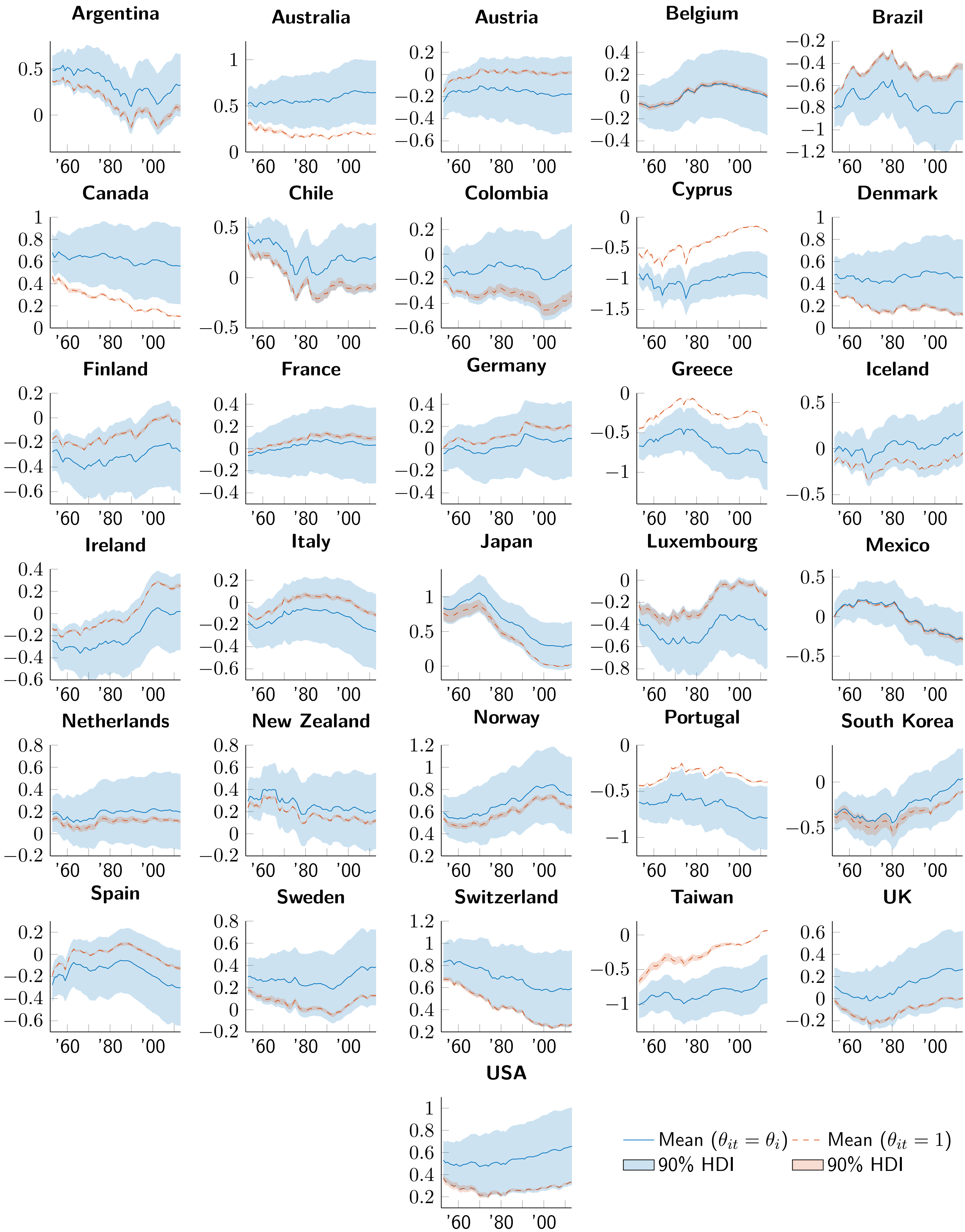
In this section we analyse the patterns for absorptive capacity revealed in our empirical results in two manners: first, we cherry-pick a number of economies on the basis of their diverging paths relative to the global frontier, and describe their policy evolution in greater detail, highlighting the correspondence with our estimated absorptive capacity evolution. Second, in order to indicate the wider validity of our results we present descriptive analysis for the full sample of (up to) 31 countries with relation to three sets of indicators highlighted in the recent Schumpeterian growth literature which dominates the current debate on policy for economic growth: aspects of financial development, tertiary education, and competition policy.

4.1 Case Studies of Structural and Economic Reforms

Ireland

Known during the 1950s as 'the poorest of the richest' economies, Ireland managed to transform its economy to one of the most productive in Europe today. Figure 3 shows Ireland's absorptive capacity to be stable until the early 1970s. This, however, was not a favorable positition since α_{it} was well below the sample average. Years of protectionism and introspective policy from the 1930s onwards effectively obstructed foreign capital flowing into Ireland. The Control of Manufactures Acts of 1932 and 1934, for instance, had the goal to ensure that new industries would be Irish-owned. Their abolishment in 1957 signaled a transition from a nationally-controlled to an outward-looking economy, a policy stance which eventually resulted in the accession to the EEC in 1973. Opening up borders for freer trade and the benefits of EEC membership led to a first surge of α_{it} during the 1970s. Seeking to boost domestic demand even further Ireland's administration turned to Keynesian

Figure 3: Posterior results α_{it}



expensionary policies. This however did not lead to the expected outcome since a substantial share of the fiscal stimulus was spent on imports, resulting in a large negative trade balance and inflationary pressure. The adverse effects on investments translated in a stagnant α_{it} during the 1980s. By the early 1990s Ireland entered a period of stunning growth in absorptive capacity. A combination of low tax rates, capital grants, a well-educated workforce and active targetting successfully attracted US high-tech companies searching for a European base. The resulting stream of incoming FDI fostered Ireland's stock of knowledge, in turn led to a steep increase of its absorptive capacity.

Sweden

Sweden's absorptive capacity evolution is charactarized by a moderate detoriation from an advantageous starting point in the 1960s. Having perhaps even fallen behind the sample average in the early 1990s the country was able to regain lost ground in just over a decade and a half. Unharmed by the widespread destruction of WWII the post-war adoption of new technologies led to the creation of a strong industrial economy, based on modern-day giants such as Volvo, Saab and Ikea which were all founded during this period. At the same time the welfare state was expanded, wage policy with centralised negotiations came into play and a higher degree of regulation applied to capital and labour markets. This evolved into a situation where the government played a pro-active role in shaping economic development and the industrial sector was strongly assisted by public investment. While this was effective in stimulating traditional manufacturing, it proved to be less fruitful during the breaktrough of microelectronics. Instead of transforming the economy throughout the 1970s and 1980s, the focus of successive goverments was to save failing industries with excessive subsidies. This hampered the incentive to develop or adopt new technologies, leading to a gradual decline of Sweden's absorptive capacity. Steps towards deregulating capital markets, enhancing competition, opening up borders even further and putting a halt to the expansion of the goverment helped to ameliorate α_{it}. Paradoxically, deregulation of capital markets brought Sweden into a financial crisis, though the resulting real economy downturn was contained efficiently by 1993. Market competition was further intensified following Sweden's accession to the EU in 1995. The outcome of these policy interventions was a clear improvement of the country's absorptive capacity since the late 1990s.

Japan

At the start of the 1950s the absorptive capacity of Japan was among the highest of all countries in our dataset and it continued to improve throughout the 1960s and 1970s. An important factor in the post-war 'Japanese miracle' was rationalization by adopting and adapting the latest vintages of foreign technology. It was the desire of the Japanese government to allocate its resources to a liminted number of industries in which it believed to possess a comparative advantage rather than allow for a market-based orientation. To this end the goverment created a number of financial intermediaries whose main task was to channel funds to key industries. On the downside, small businesses and services faced a lack of investment. Along with weak domestic competition this created a productivity disparity between these firms and the sectors targeted by the government. All in all, this strategy brought about an outward-looking economy well-equipped to incorporate technological advances. The importance of exports as an incentive to innovate cannot be underestimated, as international competition countered the disadvantages linked to weaker domestic competition. From the 1970s onwards and through the 1980s and 1990s Japan's relative absorptive capacity continuously declined, highlighting the catch-up process in manfuacturing technology in Europe and North America, fuelled in part by the widespread adoption of 'Japanese management techniques'. The model upon which Japan's success was built however appeared to be ill-suited to transform the economy towards a new reality where non-tradables and services have come to dominate. Targeting industries, protecting domestic markets, low levels of competition and excessive regulations hindered productivity growth in these markets. ICT only gradually found its way to Japanese firms as high job security made it difficult for companies to shed unskilled labour. Moreover, Japan is facing an ageing working force further holding down productivity growth.

4.2 Wider empirical evidence for structural reforms

Financial Development

A vast branch of the economics literature has successfully documented a positive link between a well-developed financial sector and economic growth through capital accumulation and technological progress. In the absence of financial intermediaries informational asymmetries, transaction costs and liquidity risk can impede an optimal allocation of capital such that innovative projects with potentionally high returns struggle to find financing (see Levine, 1997, for an in-depth discussion). Well functioning banks are able to screen new projects at lower costs and diversify risk better, making it easier to fund those start-ups with the best chances of implementing innovative products and production processes. This in turn stimulates technological progress. Theoretical evidence for a positive link between financial development and technological progress can be found *inter alia* in the endogenous growth models of De la Fuente and Marín (1996) and more recently Laeven et al. (2015). Empirically, King and Levine (1993) confirm the theory that financial services enhance growth by both fostering capital formation and improving the efficiency of that capital stock. The work by Beck et al. (2016) points to the positive association between financial innovation and capital allocation efficiency and economic growth.[6] Hsu et al. (2014) show the importance of the source of funding and that innovation in high-tech industries benefits more from equity funding as opposed to credit

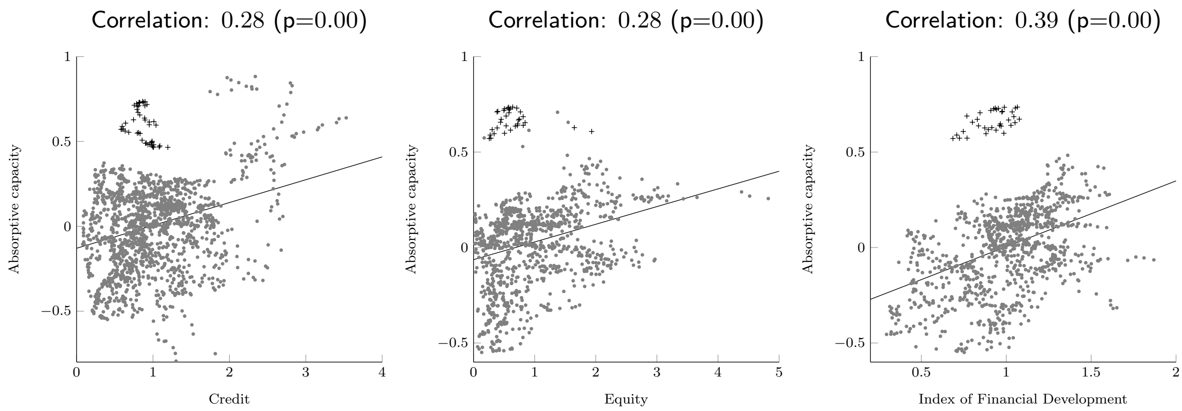
[6]They further show that financial innovation is linked to a higher appetite for risk, making bank profits more volatile, thus leading to higher losses when a banking crisis occurs. The net effect of financial intermediation, however, is positive.

funding.

Financial intermediaries lower the costs for enterpreneurs searching capital for projects that implement new production processes or the development of new products. Therefore, a lack of financing can be considered as a barrier to adoption. Figure 4 depicts scatter plots of absorptive capacity and three alternative measures for financial development, taken from the World Bank Global Financial Development Database (2016). All three measures associate a higher level of financial development with higher absorptive capacity. The first two panels plot relative absorptive capacity against private credit over GDP (credit) and stock market capitalization over GDP (equity), respectively. Given its complexity, financial deepening has several dimensions, something raw proxies such as credit or equity may not cover. To overcome this Svirydzenka (2016) created a summary index of financial development taking into account a broader range of determinants, which is shown in the third panel of Figure 4. Here and in the following two sub-sections we normalise the 'explanatory' variable (financial development, tertiary education, competition policy), such that in each year the average country score is equal to unity – this aligns the scales of these variables with that inherent in our relative absorptive capacity estimates. Finally, we highlight the results for Norway since they consistently constitute an outlier among our sample of advanced and emerging economies.

The main conclusion arising from our analysis presented in Figure 4 is that there exists a positive relationship between absorptive capacity and financial development, however, this does not seem to be particularly strong for individual measures, whereas the summary index of financial develoment provides somewhat stronger results.

Figure 4: Absorptive capacity and financial development



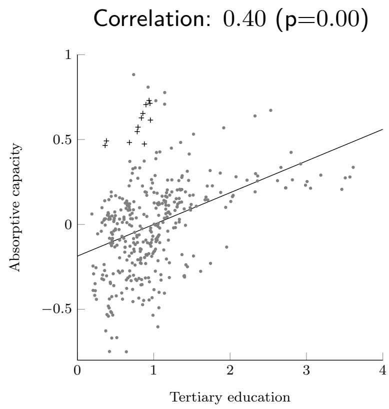
[a] Numbers in parentheses refer to p-values of the correlation coefficients.
[b] Data are normalised such that the cross-country average is one in every time period.
[c] Plus signs refer to Norway.

Human capital

The study of human capital in its (causal) relation to economic growth and development has long suffered from a failure to distinguish between the types of knowledge/education 'appropriate' at different levels of development — e.g. the Bils and Klenow (2000) 'puzzle' of comparatively low importance of education for growth; or Prichett's seminal work 'Where has all the education gone?' (Pritchett, 2001). A new consensus has recently emerged whereby tertiary-level education is seen as more relevant for countries near the technology frontier, whereas primary and secondary education are more relevant for countries far behind the frontier (Aghion and Griffith, 2005; Aghion, 2017). On balance the countries represented in our sample are those at or approaching the global technology frontier and we therefore concentrate on the link between absorptive capacity and tertiary education attainment (Aghion and Akcigit, 2017).

Our data for this exercise are taken from the standard Barro and Lee (2013) dataset for educational attainment, namely the share of population aged 25 and above who have completed tertiary education. This indicator is available for all sample countries over the 1955-2010 time horizon at 5-year intervals. Given that we adopt an attainment indicator for the entire population, the human capital aspect preferred here is that of a stock variable, rather than a flow (e.g. investment in education).

Figure 5: Absorptive capacity and tertiary education attainment



[a] See Figure 4 for details.

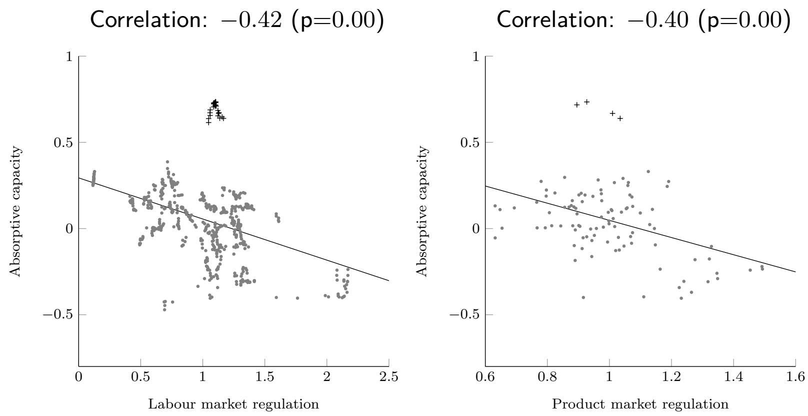
The results presented in Figure 5 indicate a strong positive correlation between relative absorptive capacity and higher educational attainment. It is notable that while the observations for Norway are on the fringes of the scatter plot they do not represent outliers to the same extent as in the financial development (and competition policy) analysis.

Competition policy

Much of the recent literature on innovation and growth has worked towards solving the often contradictory theoretical and empirical results on the role of competition by taking a more differentiated view of ‘pre-innovation’ and ‘post-innovation’ rents (Aghion and Griffith, 2005). The well-known inverted-U shape result of Aghion et al. (2005) for the competition-growth relationship is the result of a (positive) escape competition and a (negative) rent-dissipation effect with the relative magnitudes determined by the technological characteristics of the sector.

We investigate two standard measures of competition policy related to labour and product market regulation, respectively: first, employment protection legislation, measuring the costs and procedures related to dismissing individual or groups of worker(s) employed with regular contracts. These data cover 1990-2015 but are not available for Cyprus and Taiwan, and further are limited to a small number of observations in the 2010s for six primarily emerging economies. Second, product market regulation, measuring the extent to which policies inhibit or promote competition in areas of the product market where competition is viable. These data are only available for 1998, 2003, 2008 and 2013, and not at all in Argentina, Brazil, Colombia, Cyprus, and Taiwan. Both measures are collected by the OECD (2016) and associate a higher index number with more restrictive policy.

Figure 6: Absorptive capacity and competition policy indicators



[a] see Figure 4

Both scatter plots provide robust negative correlations between absorptive capacity and restrictive regulation, whether we use the large labour market regulation data in the left panel or the much scarcer product market regulation data in the right panel.

5 Summary and Conclusion

This paper introduced indeces for time-varying absorptive capacity, derived from flexible cross-country production functions estimated via Bayesian methods. Our contributions relate to (i) the econometric literature in form of an extension to the Pesaran (2006) common correlated effects (CCE) estimators to a setup where factor loadings are allowed to differ over time, a characteristic we test for as part of our implementation; and to (ii) the empirical literature on growth and productivity which to date has operationalised absorptive capacity by adopting proxies such as R&D investments or human capital, while further specifying explicit channels such as trade, FDI or migration, through which global technology can transfer to individual countries.

We estimate our model using a panel of 31 advanced economies covering 1953-2014 and present four general findings from our analysis. First, we establish that time-variation in absorptive capacity matters – failure to rejected time-invariance would have implied that our methodological contribution was superfluous, at least for the present sample and application. Absorptive capacity has changed over time, particularly so in a number of high-growth late developers including Ireland, South Korea and Taiwan. Second, we establish that for the vast majority of countries in the sample the growth boost from improvements in absorptive capacity is a one-off and does not extend into perpetuity: absorptive capacity growth (and implicitly policies which foster this growth) has TFP *levels* but not growth effects, a finding in line with theoretical models presented in Klenow and Rodríguez-Clare (2005). Third, we identify a secular process of decline in global TFP evolution, from a high and stable 2% per annum up to the late 1960s to less than 1% per annum since the mid-1990s. The period covering the Global Financial Crisis and its aftermath is too recent to allow for any meaningful prediction about the current trend in productivity: TFP growth may yet return to the stable pre-crisis mean, may still be on a declining trajectory, or may stabilise around a new level of almost zero growth. Fourth, we have employed selected country case studies as well as full sample correlation exercises to highlight the close relationship between our country- and time-specific absorptive capacity estimates and the extent of economic policy reform related to financial development, tertiary education attainment and labour and product market regulation.

The empirical analysis in this study represents merely a starting point. Our methodological contribution allows for a much richer empirical framework where we can introduce measured inputs in the innovation process (such as R&D stocks or expenditures depending on the specification) alongside the current factor error structure capturing other intangible aspects of productivity and development – this exercise could provide an investigation in parallel to the seminal Coe and Helpmann (1995) approach which still dominates parts of the literature on knowledge spillovers. We can further expand the sample of countries to move away from a focus on countries at the technology frontier and toward a study of the current 'laggards' of economic development: the analysis of absorptive capacity evolution in low- and middle-income countries can provide important insights into the differential policy implications at different levels of development. Especially in low-income countries investment in R&D is almost negligible and the estimated absorptive capacity indices enable us to identify successful countries and/or time periods which in turn can help point to suitable economic policy. Last but not least, the analysis could move away from aggregate economy data and embrace the rich sector-level data in manufacturing for advanced economies (explored in among others Griffith et al., 2004; Eberhardt et al., 2013) and in agriculture for poor and emerging economies (e.g. Eberhardt and Teal, 2013; Eberhardt and Vollrath, 2016).

References

Abramovitz, M. (1956). *Resource and Output Trends in the United States Since 1870*. National Bureau of Economic Research, Inc.

Abramovitz, M. (1986). Catching up, forging ahead, and falling behind. *The Journal of Economic History*, 46(2):385–406.

Acharya, R. C. (2016). Ict use and total factor productivity growth: intangible capital or productive externalities? *Oxford Economic Papers*, 68(1):16–39.

Acharya, R. C. and Keller, W. (2009). Technology transfer through imports. *Canadian Journal of Economics*, 42(4):1411–1448.

Aghion, P. (2017). Entrepreneurship and growth: lessons from an intellectual journey. *Small Business Economics*, 48(1):9–24.

Aghion, P. and Akcigit, U. (2017). Innovation and Growth: The Schumpeterian Perspective. In Matyas, L., editor, *Economics without Borders*, chapter 1, pages 29–72. Cambridge University Press.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, 120(2):701–728.

Aghion, P. and Griffith, R. (2005). *Competition and Growth: Reconciling Theory and Evidence*. Cambridge: MIT Press.

Aghion, P. and Howitt, P. (1998). *Endogenous Growth Theory*. MIT Press.

Aghion, P. and Jaravel, X. (2015). Knowledge spillovers, innovation and growth. *The Economic Journal*, 125(583):533–573.

Alfaro, L., Kalemli-Ozcan, S., and Volosovych, V. (2008). Why Doesn't Capital Flow from Rich to Poor Countries? An Empirical Investigation. *The Review of Economics and Statistics*, 90(2):347–368.

Andersen, T. B. and Dalgaard, C.-J. (2011). Flows of People, Flows of Ideas, and the Inequality of Nations. *Journal of Economic Growth*, 16(1):1–32.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, NBER Chapters, pages 609–626. National Bureau of Economic Research, Inc.

Bai, J., Kao, C., and Ng, S. (2009). Panel Cointegration with Global Stochastic Trends. *Journal of Econometrics*, 149(1):82–99.

Barro, R. J. and Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104:184–198.

Beck, T., Chen, T., Lin, C., and Song, F. M. (2016). Financial Innovation: The Bright and the Dark Sides. *Journal of Banking & Finance*, 72:28–51.

Bils, M. and Klenow, P. J. (2000). Does schooling cause growth? *American Economic Review*, 90(5):1160–1183.

Blanchard, O. (2004). The economic future of europe. *Journal of Economic Perspectives*, 18(4):3–26.

Bloom, N., Lemos, R., Sadun, R., Scur, D., and Van Reenen, J. (2016). International data on measuring management practices. *American Economic Review: Papers & Proceedings*, 106(5):152–156.

Carter, C. K. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81(3):541–553.

Coe, D. T. and Helpmann, E. (1995). International R&D Spillovers. *European Economic Review*, 39(5):859–887.

Cohen, W. M. and Levinthal, D. A. (1989). Innovation and Learning: The Two Faces of R&D. *The Economic Journal*, 99(397):569–596.

Corrado, C., Haskel, J., and Jona-Lasinio, C. (2017). Knowledge Spillovers, ICT and Productivity Growth. *Oxford Bulletin of Economics and Statistics*, 79(4):592–618.

De la Fuente, A. and Marín, J. M. (1996). Innovation, Bank Monitoring, and Endogenous Financial Development. *Journal of Monetary Economics*, 38(2):269–301.

Eaton, J. and Kortum, S. (1994). International Patenting and Technology Diffusion. NBER Working Papers 4931, National Bureau of Economic Research, Inc.

Eaton, J. and Kortum, S. (1999). International Technology Diffusion: Theory and Measurement. *International Economic Review*, 40(3):537–70.

Eberhardt, M., Helmers, C., and Strauss, H. (2013). Do Spillovers Matter When Estimating Private Returns to R&D? *The Review of Economics and Statistics*, 95(2):436–448.

Eberhardt, M. and Teal, F. (2013). No Mangoes in the Tundra: Spatial Heterogeneity in Agricultural Productivity. *Oxford Bulletin of Economics and Statistics*, 75(6):914–939.

Eberhardt, M. and Vollrath, D. (2016). The Role of Crop Type in Cross-Country Income Differences. CEPR Discussion Paper Series 11248.

Elliott, G. and Timmermann, A. (2005). Optimal forecast combination under regime switching*. *International Economic Review*, 46(4):1081–1102.

Ertur, C. and Musolesi, A. (2017). Weak and strong cross-sectional dependence: A panel data analysis of international technology diffusion. *Journal of Applied Econometrics*, 32(3):477–503.

Evans, P. (1997). How fast do economics converge? *The Review of Economics and Statistics*, 79(2):219–225.

Everaert, G., Heylen, F., and Schoonackers, R. (2014). Fiscal policy and tfp in the oecd: measuring direct and indirect effects. *Empirical Economics*, 49(2):605–640.

Fagerberg, J., Srholec, M., and Verspagen, B. (2009). Innovation and Economic Development. Working Papers on Innovation Studies 2009-723, Centre for Technology, Innovation and Culture, University of Oslo.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the penn world table. *American Economic Review*, 105(10):3150–82.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic Model Specification Search for Gaussian and Partial Non-Gaussian State Space Models. *Journal of Econometrics*, 154:85–100.

Gollin, D. (2002). Getting income shares right. *Journal of Political Economy*, 110(2):458–474.

Griffith, R., Redding, S., and Reenen, J. V. (2004). Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries. *The Review of Economics and Statistics*, 86(4):883–895.

Griffith, R., Redding, S., and Van Reenen, J. (2003). R&d and absorptive capacity: theory and empirical evidence. *The Scandinavian Journal of Economics*, 105(1):99–118.

Haskel, J., Corrado, C., Jona-Lasinio, C., and Iommi, M. (2013). Innovation and intangible investment in Europe, Japan and the US. Working Papers 11139, Imperial College, London, Imperial College Business School.

He, Y. and Maskus, K. E. (2012). Southern innovation and reverse knowledge spillovers: A dynamic fdi model. *International Economic Review*, 53(1):279–302.

Hsu, P.-H., Tian, X., and Xu, Y. (2014). Financial Development and Innovation: Cross-Country Evidence. *Journal of Financial Economics*, 112(1):116–135.

Kastner, G. and Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76:408 – 423.

Keller, W. (2004). International Technology Diffusion. *Journal of Economic Literature*, 42(3):752–782.

Keller, W. (2010). International Trade, Foreign Direct Investment, and Technology Spillovers. In Hall, B. H. and Rosenberg, N., editors, *Handbook of the Economics of Innovation*, volume 2, chapter 19, pages 793 – 829. North-Holland.

King, R. G. and Levine, R. (1993). Finance and Growth: Schumpeter Might Be Right. *The Quarterly Journal of Economics*, 108(3):717–737.

Klenow, P. J. and Rodríguez-Clare, A. (2005). Chapter 11 Externalities and Growth. In Aghion, P. and Durlauf, S. N., editors, *Handbook of Economic Growth*, volume 1, pages 817 – 861. Elsevier.

Koopman, S. J. and Durbin, J. (2000). Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis*, 21(3):281–296.

Kose, M. A., Otrok, C., and Prasad, E. (2012). Global business cycles: Convergence or decoupling? *International Economic Review*, 53(2):511–538.

Laeven, L., Levine, R., and Michalopoulos, S. (2015). Financial Innovation and Endogenous Growth. *Journal of Financial Intermediation*, 24(1):1–24.

Lee, K., Pesaran, M. H., and Smith, R. (1997). Growth and convergence in a multi-country empirical stochastic solow model. *Journal of Applied Econometrics*, 12(4):357–392.

Levine, R. (1997). Financial Development and Economic Growth: Views and A genda. *Journal of Economic Literature*, 35(2):688–726.

Madsen, J. B. (2007). Technology Spillover through Trade and TFP Convergence: 135 Years of Evidence for the OECD Countries. *Journal of International Economics*, 72(2):464–480.

Madsen, J. B. (2008). Economic Growth, TFP Convergence and the World Export of Ideas: A Century of Evidence. *The Scandinavian Journal of Economics*, 110(1):145–167.

Madsen, J. B., Islam, M., Ang, J. B., et al. (2010). Catching up to the technology frontier: the dichotomy between innovation and imitation. *Canadian Journal of Economics*, 43(4):1389–1411.

Miyamoto, W. and Nguyen, T. L. (2017). Business cycles in small open economies: Evidence from panel data between 1900 and 2013. *International Economic Review*, 58(3):1007–1044.

Nelson, R. R. (1959). The Economics of Invention: A Survey of the Literature. *The Journal of Business*, 32:101.

Parente, S. and Prescott, E. (1994). Barriers to Technology Adoption and Development. *Journal of Political Economy*, 102:298–321.

Parente, S. and Prescott, E. (2002). *Barriers to Riches*. MIT Press Books.

Pesaran, H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica*, 74(4):967–1012.

Pottelsberghe, B. V. and Lichtenberg, F. (2001). Does Foreign Direct Investment Transfer Technology Across Borders? *The Review of Economics and Statistics*, 83(3):490–497.

Pritchett, L. (2001). Where has all the education gone? *The World Bank Economic Review*, 15(3):367–391.

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

Strulik, H. (2014). Knowledge and growth in the very long run. *International Economic Review*, 55(2):459–482.

Svirydzenka, K. (2016). Introducing a New Broad-based Index of Financial Development. IMF Working Papers 16/5, International Monetary Fund.

Tinbergen, J. (1942). Zur theorie der langfristigen wirtschaftsentwicklung. *Weltwirtschaftliches Archiv*, 55:511–549.

van Ark, B., O'Mahoney, M., and Timmer, M. P. (2008). The Productivity Gap between Europe and the United States: Trends and Causes. *Journal of Economic Perspectives*, 22(1):25–44.

Yu, Y. and Meng, X.-L. (2011). To center or not to center: That is not the question - an ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics*, 20(3):531–570.

Online Appendix – Not intended for publication

Appendix A MCMC algorithm

In this appendix we detail the MCMC algoritm used to estimate our model in Section 2 of the main paper. We first outline the general structure of an interweaving approach to boost sampling efficiency and next provide full details on the different building blocks.

A.1 Interweaving approach

The stochastic model specification search proposed by Frühwirth-Schnatter and Wagner (2010) relies on a non-centered parametrization (NCP) of the model in which the parameters $\boldsymbol{\phi}$ and the time-varying states $\boldsymbol{s}$ are sampled in different blocks. However, the trajectories of 10,000 draws for the parameters β, ς_α and ς_θ (using the pooled estimator) plotted in the left hand side of Figure A.1 show that this blocking structure leads to extreme slow convergence of the MCMC algorithm. Inspired by Yu and Meng (2011), we boost the sampling efficiency by interweaving the NCP with a centered parametrization (CP) of the model. The basic idea is to sample the parameters $\boldsymbol{\phi}$ twice by going back and forth between the two alternative parametrizations in each iteration of the MCMC algorithm. Yu and Meng (2011) shows that by taking advantage of the contrasting features of the NCP and NC, the interweaving strategy can outperform both in terms of sampling efficiency. Minimally, it leads to an algorithm that is better than the worst of the two but often improvements are quite substantial. Kastner and Frühwirth-Schnatter (2014), for instance, use interweaving to significantly enhance sampling efficiency in stochastic volatility models.

In this section, we show how an interweaving strategy can overcome sampling deficiency in our setting. We first present the NCP and CP of the model together with their appropriate MCMC structure. Next we outline the interweaving algorithm and show the sampling efficiency gain it achieves compared to the NCP. A detailed description of the various blocks in the interwoven MCMC algorithm is provided in Subsection A.2.

Non-Centered parametrization (NCP)
The NCP of the model is given by equation (14) together with (12)-(13) in the main paper, i.e.

$$
\begin{aligned}
y_{it} &= \alpha_{i0} + \delta_\alpha \sqrt{\varsigma_\alpha}\widetilde{\alpha}_{it} + (\theta_{i0} + \delta_\theta \sqrt{\varsigma_\theta}\widetilde{\theta}_{it})\widehat{f}_t + \beta_i k_{it} + \varepsilon_{it}, && \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2),\\
\widetilde{\alpha}_{it} &= \widetilde{\alpha}_{i,t-1} + \widetilde{\psi}_{it}^\alpha, && \widetilde{\psi}_{it}^\alpha \sim N(0,1),\\
\widetilde{\theta}_{it} &= \widetilde{\theta}_{i,t-1} + \widetilde{\psi}_{it}^\theta, && \widetilde{\psi}_{it}^\theta \sim N(0,1),
\end{aligned}
$$

with $\widetilde{\alpha}_{i0}$ and $\widetilde{\theta}_{i0} = 0$. The main features of this parametrization are that: (i) it includes only the relevant time-varying components through sampling the indicators $(\delta_\theta, \delta_\alpha)$; (ii) initial conditions $(\alpha_{i0}, \theta_{i0})$ and the standard deviations $(\sqrt{\varsigma_\alpha}, \sqrt{\varsigma_\theta})$ of the innovations to the time-varying parameters are estimated and sampled as regression coefficients; (iii) the non-centered time-varying parameters $(\widetilde{\alpha}_{it}, \widetilde{\theta}_{it})$ are sampled using the Kalman filter. The general outline of the Gibbs sampler is given by:

1. Draw the binary indicators δ_θ and δ_α to determine which time-varying components should be included in the model.
2. Draw $\alpha_{i0}, \theta_{i0}, \beta_i$ together with σ_ε^2 and, if their corresponding indicator is one, $\sqrt{\varsigma_\alpha}$ and $\sqrt{\varsigma_\theta}$. When a binary indicator is zero, the corresponding standard deviation is set to zero as well.

2.* Update the common factor as $\widehat{f}_t = \overline{y}_t - \overline{\beta}\;\overline{k}_t$.

3. Draw $\widetilde{\alpha}_{it}$ and $\widetilde{\theta}_{it}$ using the Kalman filter if their corresponding binary indicator is one. When α_{it} is selected to be constant $(\delta_\alpha = 0)$, $\widetilde{\alpha}_{it}$ is sampled from its prior distribution using equation (12) in the main paper and similarly for $\widetilde{\theta}_{it}$ (when $\delta_\theta = 0$) using equation (13).

Centered parametrization (CP)
The CP of the model is given by equation (9) together with (10)-(11) in the main paper, i.e.

$$
\begin{aligned}
y_{it} &= \alpha_{it} + \theta_{it}\widehat{f}_t + \beta_i k_{it} + \varepsilon_{it}, && \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2),\\
\alpha_{it} &= \alpha_{it-1} + \psi_{it}^\alpha, && \psi_{it}^\alpha \sim N(0, \varsigma_\alpha),\\
\theta_{it} &= \theta_{it-1} + \psi_{it}^\theta, && \psi_{it}^\theta \sim N(0, \varsigma_\theta).
\end{aligned}
$$

The centered parametrization is characterized by: (i) simultaneously sampling the centered states α_{it}, θ_{it} and the fixed parameter β_i in one block using the Kalman filter; (ii) sampling innovation variances $(\varsigma_\alpha, \varsigma_\theta)$ instead of standard deviations $(\sqrt{\varsigma_\alpha}, \sqrt{\varsigma_\theta})$ in a second block. The general outline of the Gibbs sampler is given by:

1. Draw α_{it}, θ_{it} and β_i using the Kalman filter.

1.* Update the common factor as $\widehat{f}_t = \overline{y}_t - \overline{\beta}\;\overline{k}_t$.

2. Draw $\varsigma_\alpha, \varsigma_\theta$ and σ_ε^2.

Interweaving (IW)
The idea of interweaving if to sample the parameters twice, i.e. once using the CP and a second time utilizing the NCP. The general outline of our interweaving scheme is as follows:

1. Draw α_{it}, θ_{it} and β_i using the Kalman filter based on the **CP**, where α_{it} and θ_{it} are restricted to be constant when their corresponding binary indicator is zero.

1.* Update the common factor as $\widehat{f_t} = \overline{y}_t - \overline{\beta}\ \overline{k}_t$.

2. Draw ς_α and ς_θ based on the **CP** when their corresponding binary indicator is one.

2.* Move to the **NCP** using the standardizations $\widetilde{\alpha}_{it} = \frac{\alpha_{it}-\alpha_{i0}}{\sqrt{\varsigma_\alpha}}$ and $\widetilde{\theta}_{it} = \frac{\theta_{it}-\theta_{i0}}{\sqrt{\varsigma_\theta}}$ with α_{i0} and θ_{i0} being the first values of the corresponding time-varying states. When α_{it} is selected to be constant ($\delta_\alpha = 0$), $\widetilde{\alpha}_{it}$ is sampled from its prior distribution using equation (12) in the main paper and similarly for $\widetilde{\theta}_{it}$ (when $\delta_\theta = 0$) using equation (13).

3. Draw the binary indicators δ_θ and δ_α using the **NCP**.

4. Redraw $\alpha_{i0}, \theta_{i0}, \beta_i$ together with σ_ε^2 and, if their corresponding indicator is 1, $\sqrt{\varsigma_\alpha}$ and $\sqrt{\varsigma_\theta}$ using the **NCP**. When a binary indicator is zero, the corresponding standard deviation (and variance parameter) is set to zero as well.

4.* Update the common factor as $\widehat{f_t} = \overline{y}_t - \overline{\beta}\ \overline{k}_t$.

Mean Group versus Pooled estimator
We loop over the N cross-sections to obtain the heterogeneous parameters β_i and states α_{it} and θ_{it} from which the MG estimates are obtained by averaging over cross-sections. The MG estimator for β is used to update the common factor $\widehat{f_t}$. In the model where β is homogeneous, the pooled estimator for β is used to update the common factor $\widehat{f_t}$.

Sampling efficiency: non-centered parametrization versus interweaving
Figure A.1 compares the trajectories of the draws for β, ς_α and ς_θ (using the Pooled estimator) based on the NCP with those obtained from our interweaving scheme. The increase in sampling efficiency is striking for β and ς_α. For ς_θ there is no clear advantage. Highly similar results are obtained for the MG estimator.

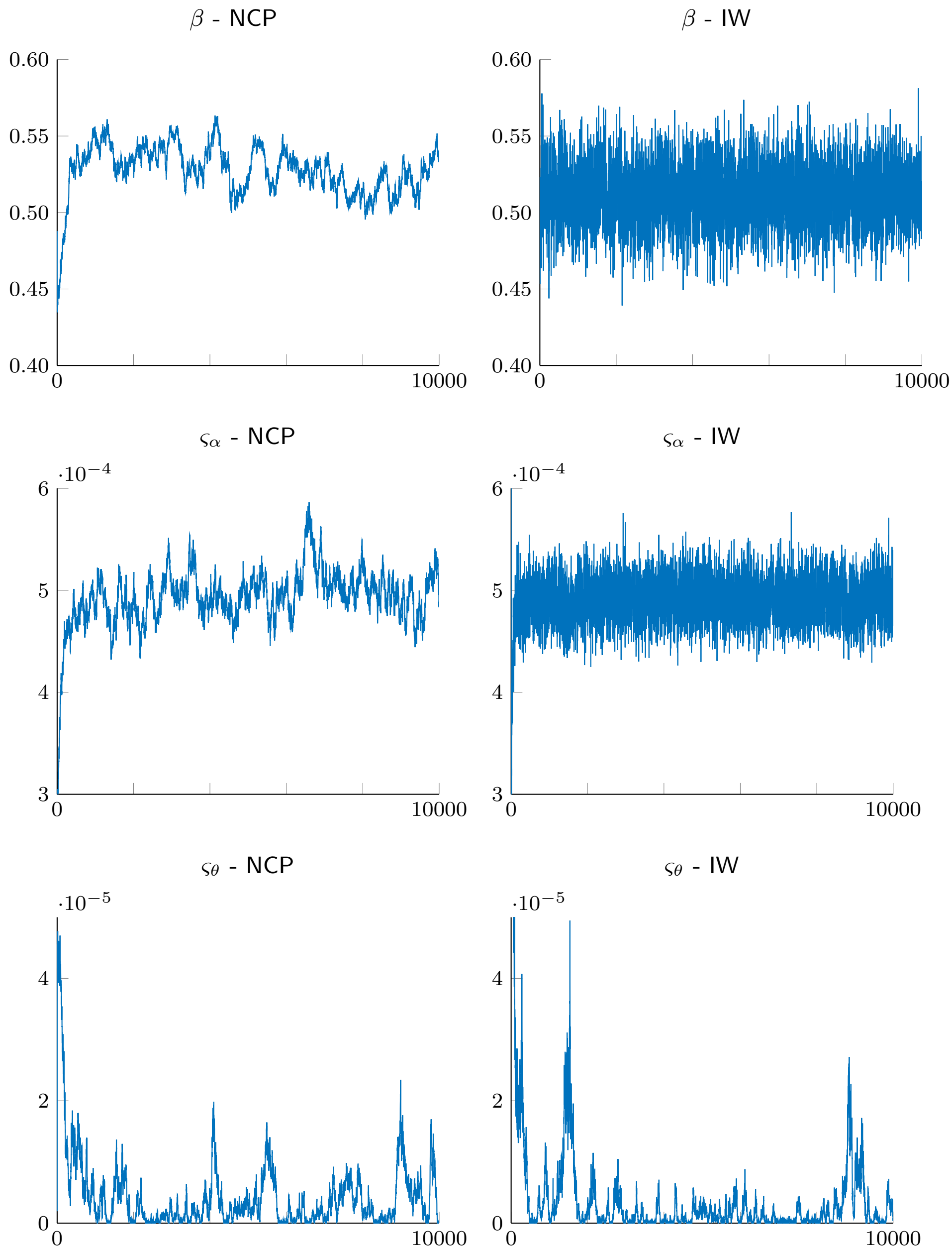
A.2 Detailed description of the interwoven MCMC algorithm

In this subsection we provide details for the four constituent blocks in the IW scheme proposed above. Steps 1.*, 2.* and 4.* should already be clear from the general outline of the IW scheme and are therefore not repeated here. Also note that our description primarily focuses on the model with heterogeneous β_i but also provides details on how the algorithm should be adjusted when β is homogeneous.

Block 1: Sampling of α_{it}, θ_{it} and β_i using the CP

In this block we sample the time-varying states α_{it} and θ_{it} together with the parameter β_i conditional on the unobserved factor $\widehat{f_t}$, the variance parameters σ_ε^2, ς_α and ς_θ and the binary indicators δ_α and δ_θ.

Figure A.1: Trajectories of the draws from the **NCP** versus the IW scheme



The conditional state space representation for cross-section i in the *heterogeneous model* is given by the observation equation

$$y_{it} = \begin{bmatrix} 1 & \widehat{f}_t & k_{it} \end{bmatrix} \begin{bmatrix} \alpha_{it} \\ \theta_{it} \\ \beta_i \end{bmatrix} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right),$$

with the evolution of the unobserved states described by

$$\begin{bmatrix} \alpha_{i,t+1} \\ \theta_{i,t+1} \\ \beta_i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{it} \\ \theta_{it} \\ \beta_i \end{bmatrix} + \begin{bmatrix} \delta_\alpha & 0 \\ 0 & \delta_\theta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \psi_{it}^\alpha \\ \psi_{it}^\theta \end{bmatrix}, \qquad \begin{bmatrix} \psi_{it}^\alpha \\ \psi_{it}^\theta \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \varsigma_\alpha & 0 \\ 0 & \varsigma_\theta) \end{bmatrix}\right).$$

The unobserved states $(\alpha_{it}, \theta_{it}, \beta_i)$ in this linear Gaussian state space model can be evaluated using the standard Kalman filter and sampled using the backward-simulation smoother of Carter and Kohn (1994). This is done for each of the N countries separately after which MG estimates can be calculated. Note that whenever a binary indicator in $(\delta_\alpha, \delta_\theta)$ equals zero, the corresponding state in $(\alpha_{it}, \theta_{it})$ is automatically restricted to be constant over time.

For the *homogeneous model*, the observation equation is given by

$$y_t = \begin{bmatrix} I_N & \widehat{f}_t I_N & k_t \end{bmatrix} \begin{bmatrix} \alpha_t \\ \theta_t \\ \beta \end{bmatrix} + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2 I_N\right),$$

where $y_t = (y_{1t}, \ldots, y_{Nt})'$, $k_t = (k_{1t}, \ldots, k_{Nt})'$ and I_N is an identity matrix of order N. The state equation now reads

$$\begin{bmatrix} \alpha_{t+1} \\ \theta_{t+1} \\ \beta \end{bmatrix} = \begin{bmatrix} I_N & 0 & 0 \\ 0 & I_N & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_t \\ \theta_t \\ \beta \end{bmatrix} + \begin{bmatrix} \delta_\alpha I_N & 0 \\ 0 & \delta_\theta I_N \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \psi_t^\alpha \\ \psi_t^\theta \end{bmatrix}, \quad \begin{bmatrix} \psi_t^\alpha \\ \psi_t^\theta \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \varsigma_\alpha I_N & 0 \\ 0 & \varsigma_\theta I_N) \end{bmatrix}\right),$$

where $\alpha_t = (\alpha_{1t}, \ldots, \alpha_{Nt})$, $\theta_t = (\theta_{1t}, \ldots, \theta_{Nt})$, $\psi_t^\alpha = (\psi_{1t}^\alpha, \ldots, \psi_{Nt}^\alpha)'$ and $\psi_t^\theta = (\psi_{1t}^\theta, \ldots, \psi_{Nt}^\theta)'$. For computational efficiency, we use the univariate treatment of the state space model when evaluating the states (see Koopman and Durbin, 2000). Sampling is again done using the backward-simulation smoother of Carter and Kohn (1994).

Block 2: Sampling ς_α and ς_θ using the CP

In this block we sample the variance parameters ς_α and ς_θ conditional on the time-varying states α_{it} and θ_{it} drawn in Block 1. Important to note is that these variances are only sampled when their corresponding binary indicator in $(\delta_\alpha, \delta_\theta)$ is one. When an indicator is set to zero (in Block 3 below), the corresponding variance parameter is also set to zero and is not sampled here.

An important aspect of the stochastic model specification search of Frühwirth-Schnatter and Wagner (2010) is that the $\mathcal{IG}$ prior on the time-varying state innovation variances ς_α and ς_θ is replaced by a

Normal prior $\mathcal{N}(0, V_0)$ on their standard deviations $\sqrt{\varsigma_\alpha}$ and $\sqrt{\varsigma_\theta}$ in the NCP. This is to avoid that the prior biases the states α_{it} and θ_{it} towards being time-varying (see discussion in Subsection 2.3). When sampling the variances from the CP it is therefore important to use a prior that is consistent with the $\mathcal{N}$ prior on the standard deviations. Following Kastner and Frühwirth-Schnatter (2014), we use a Gamma ($\mathcal{G}$) prior $\varsigma \sim V_0\chi_1^2 = \mathcal{G}(\frac{1}{2}, 2V_0)$, defined using the shape and scale parametrization, and where V_0 is the prior variance on $\sqrt{\varsigma}$ as detailed in Subsection 3.2.[7]

Since the $\mathcal{G}$ prior is non-conjugate we rely on a MetropolisÂ–Hastings (MH) step to update $\sqrt{\varsigma}$. Following Kastner and Frühwirth-Schnatter (2014), we use the auxiliary conjugate prior $p_{\text{aux}}(\varsigma) \propto \sqrt{\varsigma}^{-1}$, which denoted the improper conjugate $\mathcal{IG}(-\frac{1}{2}, 0)$ prior, to obtain suitable conditional proposal densities $p(\varsigma)$ as

$$\varsigma_\alpha|\alpha_t \sim \mathcal{IG}(c_{NT}, C_T^\alpha), \qquad \varsigma_\theta|\theta_t \sim \mathcal{IG}(c_{NT}, C_T^\theta), \tag{A.1}$$

where $c_{NT} = NT/2$, $C_T^\alpha = (\Delta\alpha_t'\Delta\alpha_t)/2$ and $C_T^\theta = (\Delta\theta_t'\Delta\theta_t)/2$. A candidate draw ς_{new} from these proposal densities is accepted with a probability of $\min(1, R)$, where

$$R = \frac{p(\varsigma_{\text{new}})}{p(\varsigma_{\text{old}})} \times \frac{p_{\text{aux}}(\varsigma_{\text{old}})}{p_{\text{aux}}(\varsigma_{\text{new}})} = \exp\left\{\frac{\varsigma_{\text{old}} - \varsigma_{\text{new}}}{2V_0}\right\}, \tag{A.2}$$

with ς_{old} denoting the last available draw for ς in the Markov chain.

Block 3: Sampling the binary indicators δ_α and δ_θ using the NCP

In this block we sample the binary indicators δ_α and δ_θ to select whether α_{it} and θ_{it} are time-varying or not. Following Frühwirth-Schnatter and Wagner (2010), when sampling these indicators we marginalize over the parameters for which variable selection is carried out. To this end, conditional on the state processes $\widetilde{\alpha}_{it}$ and $\widetilde{\theta}_{it}$, the common factor $\widehat{f}_t$ and the parameters β_i, the NCP can be written as a standard linear regression model

$$z = x^\delta b^\delta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 I_{NT}), \tag{A.3}$$

where $z = (z_1, \dots, z_N)'$, with $z_i = (z_{i1}, \dots, z_{iT})'$ and $z_{it} = y_{it} - \beta_i k_{it}$ in the *heterogeneous specification* or $z_{it} = y_{it} - \beta k_{it}$ in the *homogeneous specification*; $x = (I_N \otimes \iota_T, \widetilde{\alpha}, I_N \otimes \widehat{f}, \iota_T \otimes \widehat{f} \odot \widetilde{\theta})$, with ι_T a $(T \times 1)$ vector of ones and $\widetilde{\alpha}$ and $\widetilde{\theta}$ the time-varying parameters $\widetilde{\alpha}_{it}$ and $\widetilde{\theta}_{it}$ stacked over time and countries; $b = (\alpha_0', \sqrt{\varsigma_\alpha}, \theta_0', \sqrt{\varsigma_\theta})'$ with α_0 and θ_0 the time-invariant parameters $\widetilde{\alpha}_{i0}$ and $\widetilde{\theta}_{i0}$ stacked over countries. The vectors x^δ and b^δ exclude those elements for which the corresponding indicator in $\delta = (\delta_\alpha, \delta_\theta)$ is zero, e.g. $\widetilde{\alpha}$ is excluded from x^δ and $\sqrt{\varsigma_\alpha}$ from b^δ if $\delta_\alpha = 0$.

A naive implementation of the Gibbs sampler would be to sample δ from $g(\delta|b, z, x)$ and b from $g(b|\delta, z, x)$. Unfortunately, this approach violates conditions necessary for convergence as whenever an indicator in δ equals zero, the corresponding parameter in b is also zero which implies that the Markov chain has absorbing states. A suggested by (Frühwirth-Schnatter and Wagner, 2010), this can be avoided by marginalizing over the coefficients in b when sampling δ and subsequently drawing

[7]This is based on the general result that $X \sim N(0, \sigma^2)$ implies $X^2 \sim \sigma^2\chi_1^2 = \mathcal{G}(\frac{1}{2}, 2\sigma^2)$.

the parameters b conditional on the sampled indicators. The posterior density $g(\delta|z,x)$ can be obtained from using Bayes' Theorem as

$$g(\delta|z,x) \propto g(z|\delta,x)p(\delta), \tag{A.4}$$

where $p(\delta)$ is the prior probability of the indicators being one and $g(z|\delta,x)$ is the marginal likelihood of the regression model (A.3) where the effect of b has been integrated out. Under the the conjugate Normal-Inverse Gamma prior

$$b^{\delta} \sim \mathcal{N}(a_0^{\delta}, A_0^{\delta}\sigma_{\varepsilon}^2), \qquad \sigma_{\varepsilon}^2 \sim \mathcal{IG}(c_0, C_0), \tag{A.5}$$

the closed-form solution for $g(z|\delta,x)$ is given by

$$g(z|\delta,x) \propto \frac{\left|A_T^{\delta}\right|^{0.5}}{\left|A_0^{\delta}\right|^{0.5}} \frac{\Gamma(c_{NT})C_0^{c_0}}{\Gamma(c_0)\left(C_T^{\delta}\right)^{c_{NT}}}, \tag{A.6}$$

with posterior moments calculated as

$$a_T^{\delta} = A_T^{\delta}\left((x^{\delta})'z + (A_0^{\delta})^{-1}a_0^{\delta}\right), \tag{A.7}$$

$$A_T^{\delta} = \left((x^{\delta})'x^{\delta} + (A_0^{\delta})^{-1}\right)^{-1}, \tag{A.8}$$

$$c_{NT} = c_0 + NT/2, \tag{A.9}$$

$$C_T^{\delta} = C_0 + 0.5\left(z'z + (a_0^{\delta})'(A_0^{\delta})^{-1}a_0^{\delta} - (a_T^{\delta})'(A_T^{\delta})^{-1}a_T^{\delta}\right). \tag{A.10}$$

Instead of sampling the indicators in δ simultaneously using a multi-move sampler, we draw δ^{α} and δ^{θ} recursively from $g(\delta^{\alpha}|\delta^{\theta},z,x)$ and $g(\delta^{\theta}|\delta^{\alpha},z,x)$ using a single-move sampler where we randomize over the order in which the indicators are drawn. More specifically, the binary indicators are sampled from the Bernoulli distribution with probability

$$p(\delta^{\alpha}=1|\delta^{\theta},z,x) = \frac{g(\delta^{\alpha}=1|\delta^{\theta},z,x)}{g(\delta^{\alpha}=0|\delta^{\theta},z,x)+g(\delta^{\alpha}=1|\delta^{\theta},z,x)}, \tag{A.11}$$

and

$$p(\delta^{\theta}=1|\delta^{\alpha},z,x) = \frac{g(\delta^{\theta}=1|\delta^{\alpha},z,x)}{g(\delta^{\theta}=0|\delta^{\alpha},z,x)+g(\delta^{\theta}=1|\delta^{\alpha},z,x)}. \tag{A.12}$$

Block 4: Sampling the parameters α_{i0}, θ_{i0}, $\sqrt{\varsigma_\alpha}$, $\sqrt{\varsigma_\theta}$, β_i and σ_ε^2 using the NCP

In this last step we sample the variance σ_{ε}^2 of the observation errors from $\mathcal{IG}(c_T, C_T^{\delta})$ and the (unrestricted) parameters in $b = (\alpha_0,\ \sqrt{\varsigma_\alpha},\ \theta_0,\ \sqrt{\varsigma_\theta},\ \beta_1,\dots,\beta_N)$ from $\mathcal{N}(a_T^{\delta}, A_T^{\delta}\sigma_{\varepsilon}^2)$, using the regression model (A.3), replacing z by y and redefining $x = (I_N \otimes \iota_T, \widetilde{\alpha}, I_N \otimes \widehat{f}, \iota_T \otimes \widehat{f} \odot \widetilde{\theta}, \text{diag}(k_1,\dots,k_N))$ with $k_i = (k_{i1},\dots,k_{iT})$. In the *homogeneous model* we set $x = (I_N \otimes \iota_T, \widetilde{\alpha}, I_N \otimes \iota_T, \widetilde{\theta}, k)$ and $b = (\alpha_0,\ \sqrt{\varsigma_\alpha},\ \theta_0,\ \sqrt{\varsigma_\theta},\ \beta)$.

When a binary indicator in δ is zero, the corresponding variance parameter is not sampled but restricted to be zero. To re-enforce the fact that the sign of the standard deviations $(\sqrt{\varsigma_\alpha}, \sqrt{\varsigma_\alpha})$ and the states $(\widetilde{\alpha}_{it}, \widetilde{\theta}_{it})$ are not separately identified, we perform a random sign switch, e.g. $\sqrt{\varsigma_\alpha}$ and $\widetilde{\alpha}_{it}$ are left unchanged with probability 0.5 while with the same probability they are replaced by $-\sqrt{\varsigma_\alpha}$ and $-\widetilde{\alpha}_{it}$.

Appendix B Additional Empirical Results

Figure B.1: Posterior results $\widehat{f}_t$ – MG specification

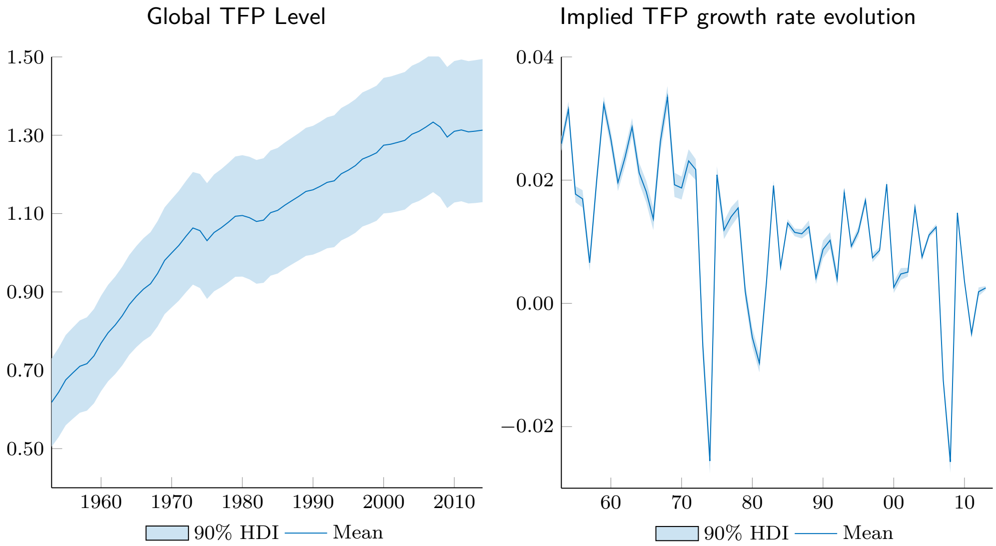


Table B.1: Posterior results for β_i (Unrestricted MG)

	β_i			β_i	
Argentina	0.34	(0.08)	Italy	0.60	(0.08)
Australia	0.69	(0.10)	Japan	0.34	(0.04)
Austria	0.63	(0.08)	Luxembourg	0.35	(0.11)
Belgium	0.65	(0.09)	Mexico	0.68	(0.10)
Brazil	0.67	(0.08)	Netherlands	0.59	(0.09)
Canada	0.51	(0.10)	NewZealand	0.72	(0.12)
Chile	0.40	(0.07)	Norway	0.67	(0.09)
Colombia	0.58	(0.13)	Portugal	0.59	(0.08)
Cyprus	0.41	(0.08)	SouthKorea	0.61	(0.04)
Denmark	0.54	(0.08)	Spain	0.60	(0.06)
Finland	0.45	(0.08)	Sweden	0.51	(0.09)
France	0.64	(0.08)	Switzerland	0.41	(0.10)
Germany	0.62	(0.08)	Taiwan	0.59	(0.05)
Greece	0.58	(0.06)	UK	0.59	(0.10)
Iceland	0.56	(0.08)	USA	0.55	(0.10)
Ireland	0.60	(0.07)			

[a] Standard deviations of the posterior distributions are reported in parentheses.
[b] Based on MCMC with $K = 45,000$ iterations where the first $B = 5,000$ are discarded as burn-in.

Appendix C Additional Figures

Figure C.1: Real R&D expenditure growth evolution (smoothened country paths)