Online Appendix — Estimating and Testing the Multicountry Endogenous Growth Model

By Stef De Visscher, Markus Eberhardt and Gerdie Everaert

Appendix A MCMC algorithm

In this appendix we detail the MCMC algorithm that is used to estimate the baseline and extended model presented in the main paper. In Subsection A.1, we first outline the general structure of an interweaving approach to boost sampling efficiency. Next, in Subsection A.2, we provide full details on the different building blocks of our MCMC algorithm. In Subsection A.3 we perform a Monte Carlo simulation to analyze small sample performance.

A.1 Interweaving approach

The stochastic model specification search proposed by Frühwirth-Schnatter and Wagner (2010) relies on a non-centered parameterization (NCP) - as in equation (14) of the main paper - in which the parameters $\boldsymbol{\phi}$ and the time-varying states $\boldsymbol{s}$ are sampled in different blocks. However, the trajectories of 20,000 draws for the parameters ς_α and ς_θ, using the MCMC algorithm on simulated data, plotted in the left column of Figure A.1, show that this parameterization leads to very slow convergence. This is confirmed by the integrated autocorrelation time, which is 986 for ς_α and 1,945 for ς_θ. The integrated autocorrelation time is calculated as $\tau = 1 + 2\sum_{t=1}^{d} \widehat{c}(t)$ with $\widehat{c}(t)$ the estimated autocorrelation function at lag t and d the lag for which $\widehat{c}(t) < 0.01$. It can be interpreted as the factor by which the squared Monte Carlo standard error increases due to the dependence in the Markov chain. The integrated autocorrelation time equals one in the ideal situation of complete independence, while higher values imply a reduction in the 'effective' number of draws such that more MCMC iterations are needed to attain the same Monte Carlo standard error.

Inspired by Yu and Meng (2011), we boost the sampling efficiency by interweaving the NCP with a centered parameterization (CP) of the model. The basic idea is to sample the parameters $\boldsymbol{\phi}$ twice by going back and forth between the two alternative parameterizations in each iteration of the MCMC algorithm. Yu and Meng (2011) shows that by taking advantage of the contrasting features of the NCP and CP, 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. A similar interweaving approach is used by Bitto and Frühwirth-Schnatter (2019) to achieve shrinkage in a time-varying parameter model framework.

In the remainder of 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. We will focus on the following extended model (as presented in Section 4.1 of the main paper):

$$\lambda_{it} = \alpha^*_{it} + \beta_{jt}x_{j,it} + \sum_{k=1,k\neq j}^{K}(\beta^*_{k\alpha} + \beta_{jt}\beta^*_{k\theta})x_{k,it} + \theta^*_{it}\overline{\lambda}_t + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \mathcal{N}(0, \sigma^2_\varepsilon), \tag{A.1}$$

$$\alpha^*_{it} = \alpha^*_{it-1} + \psi^\alpha_{it}, \qquad \psi^\alpha_{it} \sim \mathcal{N}(0, \varsigma_\alpha), \tag{A.2}$$

$$\theta^*_{it} = \theta^*_{it-1} + \psi^\theta_{it}, \qquad \psi^\theta_{it} \sim \mathcal{N}(0, \varsigma_\theta), \tag{A.3}$$

$$\beta_{jt} = \beta_{jt-1} + \psi^\beta_{jt}, \qquad \psi^\beta_{jt} \sim \mathcal{N}(0, \varsigma_\beta). \tag{A.4}$$

The restricted version in equation (24) is obtained by setting $\beta_{jt} = \beta_{j\alpha}$, $\beta^*_{k\theta} = 0$ and, $\beta^*_{k\alpha} = \beta_{k\alpha}$ $(\forall k)$. The baseline model presented in Section 2 is obtained by further setting $\beta_{k\alpha} = 0$ $(\forall k)$.

Centered parameterization (CP)

Equations (A.1)-(A.4) constitute the CP of the model. This parameterization is characterized by: (i) sampling the centered states $(\alpha^*_{it}, \theta^*_{it}, \beta_{jt})$ and $\beta^{*\backslash j}_\alpha = (\beta^*_{1\alpha}, \ldots, \beta^*_{j-1,\alpha}, \beta^*_{j+1,\alpha}, \ldots \beta^*_{K\alpha})$ using the Kalman filter; (ii) sampling innovation variances $(\varsigma_\alpha, \varsigma_\theta, \varsigma_\beta)$ instead of standard deviations $(\sqrt{\varsigma_\alpha}, \sqrt{\varsigma_\theta}, \sqrt{\varsigma_\beta})$ in a second block; (iii) sampling $\beta^{*\backslash j}_\theta = (\beta^*_{1\theta}, \ldots, \beta^*_{j-1,\theta}, \beta^*_{j+1,\theta}, \ldots \beta^*_{K\theta})$ and σ^2_ε together with σ^2_ε in a third block. The general outline of the Gibbs sampler is given by:

1. Draw α^*_{it}, θ^*_{it}, β_{jt} and $\beta^{*\backslash j}_\alpha$ using the Kalman filter.
2. Draw ς_α, ς_θ and ς_β.
3. Draw $\beta^{*\backslash j}_\theta$.

Non-Centered parameterization (NCP)

The NCP of the model in equations (A.1)-(A.3) is given by:

$$\lambda_{it} = \alpha^*_{it} + \beta_{jt}x_{j,it} + \sum_{k=1,k\neq j}^{K}(\beta^*_{k\alpha} + \beta_{jt}\beta^*_{k\theta})x_{k,it} + \theta^*_{it}\overline{\lambda}_t + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \mathcal{N}(0, \sigma^2_\varepsilon), \tag{A.5}$$

$$\alpha^*_{it} = \alpha^*_{i0} + \delta_\alpha\sqrt{\varsigma_\alpha}\widetilde{\alpha}_{it}, \quad \text{with } \widetilde{\alpha}_{it} = \widetilde{\alpha}_{it-1} + \widetilde{\psi}^\alpha_{it}, \quad \widetilde{\alpha}_{i0} = 0, \quad \widetilde{\psi}^\alpha_{it} \sim \mathcal{N}(0,1), \tag{A.6}$$

$$\theta^*_{it} = \theta^*_{i0} + \delta_\theta\sqrt{\varsigma_\theta}\widetilde{\theta}_{it}, \quad \text{with } \widetilde{\theta}_{it} = \widetilde{\theta}_{it-1} + \widetilde{\psi}^\theta_{it}, \quad \widetilde{\theta}_{i0} = 0, \quad \widetilde{\psi}^\theta_{it} \sim \mathcal{N}(0,1), \tag{A.7}$$

$$\beta_{jt} = \beta_{j0} + \delta_\beta\sqrt{\varsigma_\beta}\widetilde{\beta}_{jt}, \quad \text{with } \widetilde{\beta}_{jt} = \widetilde{\beta}_{jt-1} + \widetilde{\psi}^\beta_t, \quad \widetilde{\beta}_{j0} = 0, \quad \widetilde{\psi}^\beta_t \sim \mathcal{N}(0,1). \tag{A.8}$$

The main features of this parameterization are that: (i) it includes only the relevant time-varying components through sampling the indicators $(\delta_\alpha, \delta_\theta, \delta_\beta)$; (ii) the initial conditions $(\alpha^*_{i0}, \theta^*_{i0}, \beta_{j0})$, the standard

deviations ($\sqrt{\varsigma_\alpha}, \sqrt{\varsigma_\theta}, \sqrt{\varsigma_\beta}$) of the innovations to the time-varying parameters and the parameters in β_α^* and $\beta_\theta^{*\backslash j}$ are estimated and sampled as regression coefficients; (iii) the non-centered time-varying parameters ($\widetilde{\alpha}_{it}^*, \widetilde{\theta}_{it}^*, \widetilde{\beta}_{jt}$) are sampled using the Kalman filter. The general outline of the Gibbs sampler is given by:

1. Draw the binary indicators δ_α, δ_θ and δ_β to determine which time-varying components should be included in the model.

2. Draw α_{i0}^*, θ_{i0}^*, β_{j0}, $\beta_\alpha^{*\backslash j}$ together with σ_ε^2 and, if their corresponding indicator is one, $\sqrt{\varsigma_\alpha}$, $\sqrt{\varsigma_\theta}$ and $\sqrt{\varsigma_\beta}$. When a binary indicator is zero, the corresponding standard deviation is set to zero as well.

3. Draw $\beta_\theta^{*\backslash j}$ if $\delta_\beta = 1$. When $\delta_\beta = 0$, $\beta_\theta^{*\backslash j}$ is set to zero.

4. Draw $\widetilde{\alpha}_{it}^*$, $\widetilde{\theta}_{it}^*$ and $\widetilde{\beta}_{jt}$ 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 (A.6) and similarly for $\widetilde{\theta}_{it}^*$ (when $\delta_\theta = 0$) and $\widetilde{\beta}_{jt}$ (when $\delta_\beta = 0$) using equations (A.7) and (A.8).

Interweaving (IW)

The idea of interweaving is 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}^*, β_{jt} and $\beta_\alpha^{*\backslash j}$ using the Kalman filter based on the **CP**, where α_{it}^*, θ_{it}^* and β_{jt} are restricted to be constant when their corresponding binary indicator is zero.

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

2.* Move to the **NCP** using the standardizations $\widetilde{\alpha}_{it}^* = \frac{\alpha_{it}^* - \alpha_{i0}^*}{\sqrt{\varsigma_\alpha}}$, $\widetilde{\theta}_{it}^* = \frac{\theta_{it}^* - \theta_{i0}^*}{\sqrt{\varsigma_\theta}}$ and $\widetilde{\beta}_{jt} = \frac{\beta_{jt} - \beta_{j0}}{\sqrt{\varsigma_\beta}}$ with α_{i0}^*, θ_{i0}^* and β_{j0} 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 (A.6) and similarly for $\widetilde{\theta}_{it}^*$ (when $\delta_\theta = 0$) and $\widetilde{\beta}_{jt}$ (when $\delta_\beta = 0$) using equations (A.7) and (A.8).

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

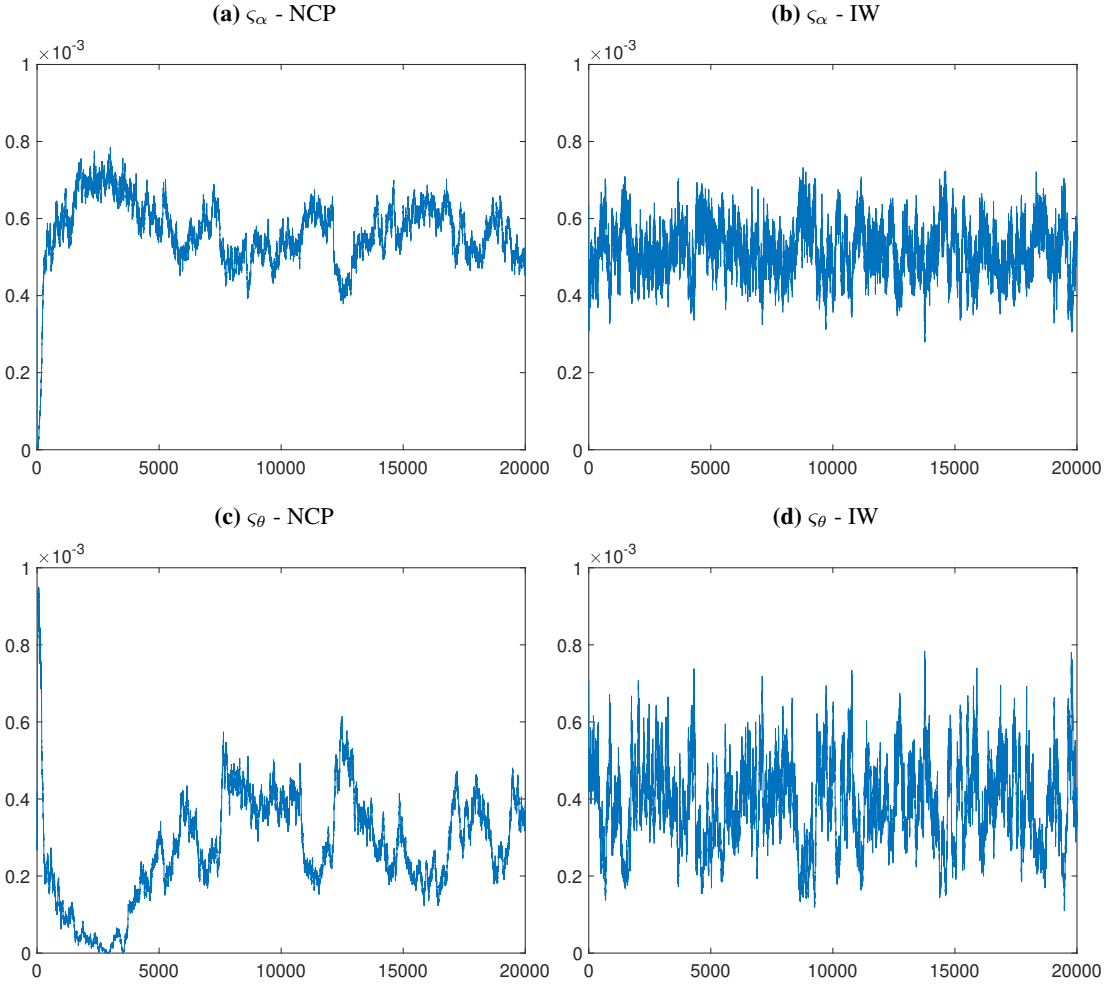
4. Redraw α_{i0}^*, θ_{i0}^* and β_{j0} together with β_α^* and σ_ε^2 and, if their corresponding indicator is 1, $\sqrt{\varsigma_\alpha}$, $\sqrt{\varsigma_\theta}$ and $\sqrt{\varsigma_\beta}$ using the **NCP**. When a binary indicator is zero, the corresponding standard deviation (and variance parameter) is set to zero as well.

5. Draw $\beta_\theta^{*\backslash j}$ if $\delta_\beta = 1$, when $\delta_\beta = 0$ set $\beta_\theta^{*\backslash j} = 0$ and continue with block 1. To improve the sampling efficiency we redraw the parameters α_{i0}^*, θ_{i0}^*, $\beta_\alpha^{*\backslash j}$ and σ_ε^2 and, if their corresponding indicator is 1, $\sqrt{\varsigma_\alpha}$ and $\sqrt{\varsigma_\theta}$.

Sampling efficiency: non-centered parameterization versus interweaving

Figure A.1 compares trajectories of draws for ς_α and ς_θ obtained from using the NCP on simulated data with those from the proposed interweaving scheme. The increase in sampling efficiency is clearly visible from the graph. More formally, the integrated autocorrelation time drops from 986 to 139 for ς_α and from 1,945 to 160 for ς_θ when moving from the NCP to the interweaving scheme.

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



Note: Based on simulated data, i.e. one draw using scenario 2 ($\delta_\alpha = \delta_\theta = 1$) from the Monte Carlo simulation design as outlined in Section A.3 below.

A.2 Detailed description of the interwoven MCMC algorithm

In this subsection we provide details for the constituent blocks of the interwoven MCMC algorithm proposed above. Step 2.* should already be clear from the general outline of the IW scheme and is therefore not repeated here.

Block 1: Sampling $\alpha^*_{it}, \theta^*_{it}, \beta_{jt}$ and $\beta^{*\setminus j}_\alpha$ using the CP

In this block we sample the time-varying states α^*_{it}, θ^*_{it} and β_{jt} together with the parameters $\beta^{*\setminus j}_\alpha$ condi-

tional on the variance parameters (ς_α, ς_θ, ς_β, σ_ε^2), the binary indicators (δ_α, δ_θ, δ_β) and $\beta_\theta^{*\backslash j}$. Define the auxiliary variable $\widetilde{x}_{it} = x_{j,it} + x_{it}^{\backslash j}\beta_\theta^*$, with $x_{it}^{\backslash j} = (x_{1,it}, \ldots, x_{j-1,it}, x_{j+1,it}, \ldots, x_{K,it})$. The conditional state-space representation is given by the observation equation

$$\lambda_t = \begin{bmatrix} I_N & \overline{\lambda}_t I_N & \widetilde{x}_t & x_t^{\backslash j} \end{bmatrix} \begin{bmatrix} \alpha_t^* \\ \theta_t^* \\ \beta_{jt} \\ \beta_\alpha^{*\backslash j} \end{bmatrix} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2 I_N\right),$$

where $\lambda_t = (\lambda_{1t}, \ldots, \lambda_{Nt})'$, $\widetilde{x}_t = (\widetilde{x}_{1t}, \ldots, \widetilde{x}_{Nt})'$, $x_t^{\backslash j} = (x_{1t}^{\backslash j}, \ldots, x_{Nt}^{\backslash j})'$, I_N is an identity matrix of order N and the evolution of the unobserved states described by

$$\begin{bmatrix} \alpha_{t+1}^* \\ \theta_{t+1}^* \\ \beta_{jt+1} \\ \beta_\alpha^{*\backslash j} \end{bmatrix} = \begin{bmatrix} I_N & 0 & 0 & 0 \\ 0 & I_N & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & I_{K-1} \end{bmatrix} \begin{bmatrix} \alpha_t^* \\ \theta_t^* \\ \beta_{jt} \\ \beta_\alpha^{*\backslash j} \end{bmatrix} + \begin{bmatrix} \delta_\alpha I_N & 0 & 0 \\ 0 & \delta_\theta I_N & 0 \\ 0 & 0 & \delta_\beta \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \psi_t^\alpha \\ \psi_t^\theta \\ \psi_t^\beta \end{bmatrix},$$

$$\begin{bmatrix} \psi_t^\alpha \\ \psi_t^\theta \\ \psi_t^\beta \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \varsigma_\alpha I_N & 0 & 0 \\ 0 & \varsigma_\theta I_N & 0 \\ 0 & 0 & \varsigma_\beta \end{bmatrix}\right),$$

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 $\psi_t^\theta = (\psi_{1t}^\theta, \ldots, \psi_{Nt}^\theta)'$. The unobserved states $(\alpha_t^*, \theta_t^*, \beta_{jt})$ 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). Note that whenever a binary indicator δ_α, δ_θ or δ_β equals zero, the corresponding states α_t^*, θ_t^* or β_{jt} are automatically restricted to be constant over time. Note that the restricted model in equation (24) of Section 4.1 - where the variables in x don't have growth effects - can be obtained by setting $\beta_\theta^{*\backslash j} = 0$, $\widetilde{x}_{it} = x_{j,it}$, $\delta_\beta = 0$ and $\beta_{jt} = \beta_j$ such that $\beta_\alpha = (\beta_j, \beta_\alpha^{*\backslash j})'$.

Block 2: Sampling ς_α, ς_θ and ς_β using the CP

In this block we sample the variance parameters ς_α, ς_θ and ς_β conditional on the time-varying states α_t^*, θ_t^* and β_{jt} drawn in Block 1. These variances are only sampled when their corresponding binary indicator δ_α, δ_θ or δ_β 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 Inverse Gamma $\mathcal{IG}$ prior on the time-varying state innovation variances ς is replaced by a Normal prior $\mathcal{N}(0, V_0)$ on their standard deviations $\sqrt{\varsigma}$ in the NCP. This is to avoid that the prior biases the states α_{it}^*, θ_{it}^* and β_{jt} towards being time-varying (see discussion in Subsection 2.3 of the main paper). 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 $\varsigma \sim V_0\chi^2_1 = \mathcal{G}(\frac{1}{2}, 2V_0)$, defined using the shape and scale parameterization, and where V_0 is the prior variance on $\sqrt{\varsigma}$ as detailed in Subsection 3.2.[1]

Since the $\mathcal{G}$ prior is non-conjugate we rely on a Metropolis–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 denotes the density of an auxiliary improper conjugate $\mathcal{IG}(-\frac{1}{2}, 0)$ prior, to obtain suitable conditional proposal densities $p(\varsigma)$ as

$$\varsigma_\alpha|\alpha^*_t \sim \mathcal{IG}(c_{NT}, C^\alpha_T), \qquad \varsigma_\theta|\theta^*_t \sim \mathcal{IG}(c_{NT}, C^\theta_T), \qquad \varsigma_\beta|\beta_{jt} \sim \mathcal{IG}(c_T, C^\theta_T) \tag{A.9}$$

where $c_{NT} = NT/2$, $c_T = T/2$, $C^\alpha_T = (\Delta\alpha^{*'}_t \Delta\alpha^*_t)/2$, $C^\theta_T = (\Delta\theta^{*'}_t \Delta\theta^*_t)/2$ and $C^\beta_T = (\Delta\beta^{*'}_t \Delta\beta^*_t)/2$. A candidate draw ς_{new} from these proposal densities is accepted with a probability of $\min(1, R)$, where

$$R = \frac{p(\varsigma_{\text{new}})}{p(\varsigma_{\text{old}})} \times \frac{p_{\text{aux}}(\varsigma_{\text{old}})}{p_{\text{aux}}(\varsigma_{\text{new}})} = \exp\left\{\frac{\varsigma_{\text{old}} - \varsigma_{\text{new}}}{2V_0}\right\}, \tag{A.10}$$

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

Block 3: Sampling the binary indicators $\boldsymbol{\delta_\alpha}$, $\boldsymbol{\delta_\theta}$ and $\boldsymbol{\delta_\beta}$ using the NCP

In this block we sample the binary indicators δ_α, δ_θ and δ_β to select whether α^*_{it}, θ^*_{it} and β_{jt} vary over time 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 ($\widetilde{\alpha}^*_{it}, \widetilde{\theta}^*_{it}, \widetilde{\beta}_{jt}$) and the parameters $\beta^{*\setminus j}_\alpha$ and $\beta^{*\setminus j}_\theta$, the NCP can be written as a standard linear regression model

$$z = w^\delta b^\delta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2_\varepsilon I_{NT}), \tag{A.11}$$

where $z = (z_1, \ldots, z_N)'$, with $z_i = (z_{i1}, \ldots, z_{iT})'$ and $z_{it} = \lambda_{it} - x^{\setminus j}_{it}\beta^*_\alpha$; $w = (I_N \otimes \iota_T, \widetilde{\alpha}^*, I_N \otimes \overline{\lambda}, \iota_N \otimes \overline{\lambda} \odot \widetilde{\theta}, \widetilde{x}, \iota_N \otimes \widetilde{\beta}_j \odot \widetilde{x})$, with ι_T a $(T \times 1)$ vector of ones, $\widetilde{\alpha}^*$ and $\widetilde{\theta}^*$ the time-varying parameters $\widetilde{\alpha}^*_{it}$ and $\widetilde{\theta}^*_{it}$ stacked over time and countries, $\widetilde{x}$ the variable $\widetilde{x}_{it}$ stacked over time and countries and $\widetilde{\beta}_j$ the time-varying parameter $\widetilde{\beta}_{jt}$ stacked over time; $b = (\alpha^{*'}_0, \sqrt{\varsigma_\alpha}, \theta^{*'}_0, \sqrt{\varsigma_\theta}, \beta_{j0}, \sqrt{\varsigma_\beta})'$ with α^*_0 and θ^*_0 the time-invariant parameters α^*_{i0} and θ^*_{i0} stacked over countries. The vectors w^δ and b^δ exclude those elements in w and b for which the corresponding indicator in $\delta = (\delta_\alpha, \delta_\theta, \delta_\beta)$ is zero, e.g. $\widetilde{\alpha}^*$ is excluded from w^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 $g(\delta|b, z, w)$ and b from

[1] This is based on the general result that $X \sim \mathcal{N}(0, \sigma^2)$ implies $X^2 \sim \sigma^2\chi^2_1 = \mathcal{G}(\frac{1}{2}, 2\sigma^2)$.

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

$$g(\delta|z, w) \propto g(z|\delta, w)p(\delta), \tag{A.12}$$

where $p(\delta)$ is the prior probability of the indicators being one and $g(z|\delta, w)$ is the marginal likelihood of the regression model (A.11) where the effect of b has been integrated out. Under the the conjugate Normal-Inverse Gamma prior

$$b^\delta \sim \mathcal{N}(a_0^\delta, A_0^\delta \sigma_\varepsilon^2), \qquad \sigma_\varepsilon^2 \sim \mathcal{IG}(c_0, C_0), \tag{A.13}$$

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

$$g(z|\delta, w) \propto \frac{\left|A_T^\delta\right|^{0.5}}{\left|A_0^\delta\right|^{0.5}} \frac{\Gamma(c_{NT}) C_0^{c_0}}{\Gamma(c_0) \left(C_T^\delta\right)^{c_{NT}}}, \tag{A.14}$$

with posterior moments calculated as

$$a_T^\delta = A_T^\delta \left((w^\delta)' z + (A_0^\delta)^{-1} a_0^\delta\right), \tag{A.15}$$

$$A_T^\delta = \left((w^\delta)' w^\delta + (A_0^\delta)^{-1}\right)^{-1}, \tag{A.16}$$

$$c_{NT} = c_0 + NT/2, \tag{A.17}$$

$$C_T^\delta = C_0 + 0.5 \left(z'z + (a_0^\delta)'(A_0^\delta)^{-1} a_0^\delta - (a_T^\delta)'(A_T^\delta)^{-1} a_T^\delta\right). \tag{A.18}$$

Instead of sampling the indicators in δ simultaneously using a multi-move sampler, we draw δ^α, δ^θ δ_β recursively from $g(\delta_\alpha|\delta_\theta, \delta_\beta, z, w)$, $g(\delta_\theta|\delta_\alpha, \delta_\beta, z, w)$ and $g(\delta_\beta|\delta_\alpha, \theta_\beta, z, w)$ 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 probabilities

$$p(\delta_\alpha = 1|\delta_\theta, \delta_\beta, z, w) = \frac{g(\delta_\alpha = 1|\delta_\theta, \delta_\beta, z, w)}{g(\delta_\alpha = 0|\delta_\theta, \delta_\beta, z, w) + g(\delta_\alpha = 1|\delta_\theta, \delta_\beta, z, w)}, \tag{A.19}$$

$$p(\delta_\theta = 1|\delta_\alpha, \delta_\beta, z, w) = \frac{g(\delta_\theta = 1|\delta_\alpha, \delta_\beta, z, w)}{g(\delta_\theta = 0|\delta_\alpha, \delta_\beta, z, w) + g(\delta_\alpha, \delta_\beta = 1|\delta^\alpha, z, w)}, \tag{A.20}$$

$$p(\delta_\beta = 1|\delta_\alpha, \delta_\theta, z, w) = \frac{g(\delta_\beta = 1|\delta_\alpha, \delta_\theta, z, w)}{g(\delta_\beta = 0|\delta_\alpha, \delta_\theta, z, w) + g(\delta_\beta = 1|\delta_\alpha, \delta_\theta, z, w)}. \tag{A.21}$$

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

In this block we sample the variance σ_ε^2 of the observation errors from $\mathcal{IG}(c_T, C_T^\delta)$ and the (unrestricted) parameters in $b = (\alpha_0^{*\prime}, \sqrt{\varsigma_\alpha}, \theta_0^{*\prime}, \sqrt{\varsigma_\theta}, \beta_{j0}, \sqrt{\varsigma_\beta}, \beta_\alpha^{*\backslash j})'$ from $\mathcal{N}(a_T^\delta, A_T^\delta \sigma_\varepsilon^2)$, using the regression model (A.11), with $z = \lambda$ and redefining $w = (I_N \otimes \iota_T, \widetilde{\alpha}^*, I_N \otimes \overline{\lambda}, \iota_N \otimes \overline{\lambda} \odot \widetilde{\theta}, \widetilde{x}, \iota_N \otimes \widetilde{\beta} \odot \widetilde{x}, x^{\backslash j})$ with $x^{\backslash j}$ the data in $x_{it}^{\backslash j}$ stacked over time and countries. 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 $(\sqrt{\varsigma_\alpha}, \sqrt{\varsigma_\alpha}, \sqrt{\varsigma_\beta})$ and the states $(\widetilde{\alpha}_{it}^*, \widetilde{\theta}_{it}^*, \widetilde{\beta}_{jt})$ are not separately identified, we perform a random sign switch, e.g. $\sqrt{\varsigma_\alpha}$ is left unchanged with probability 0.5 while with the same probability it is replaced by $-\sqrt{\varsigma_\alpha}$.

Block 5: Sampling $\beta_\theta^{*\backslash j}$

In this last step we sample $\beta_\theta^{*\backslash j}$ if δ_β is 1, when $\delta_\beta = 0$ set $\beta_\theta^{*\backslash j} = 0$ and continue with block 1. To improve the sampling efficiency we redraw the parameters $(\alpha_0^{*\prime}, \sqrt{\varsigma_\alpha}, \theta_0^{*\prime}, \sqrt{\varsigma_\theta}, \beta_\alpha^{*\backslash j}, \sigma_\varepsilon^2)$. Again, we sample the variance σ_ε^2 of the observation errors from $\mathcal{IG}(c_T, C_T^\delta)$ and the (unrestricted) parameters in $b = (\alpha_0^*, \sqrt{\varsigma_\alpha}, \theta_0^*, \sqrt{\varsigma_\theta}, \beta_\alpha^{*\backslash j}, \beta_\beta^*)$ from $\mathcal{N}(a_T^\delta, A_T^\delta \sigma_\varepsilon^2)$, using the regression model (A.11), redefining $z_{it} = \lambda_{it} - \beta_{jt} x_{j,it}$ and $w = (I_N \otimes \iota_T, \widetilde{\alpha}^*, I_N \otimes \widehat{f}, \iota_T \otimes \overline{\lambda} \odot \widetilde{\theta}^*, x^{\backslash j}, \iota_N \iota_{K-1}{}' \otimes \beta_j \odot x^{\backslash j})$ where β_j is β_{jt} stacked over time. When a binary indicator in $(\delta_\alpha, \delta_\theta)$ is zero, the corresponding variance parameter is not sampled but restricted to be zero.

A.3 Monte Carlo simulation

In this section, we present Monte Carlo simulation results to examine the performance of the MCMC algorithm outlined in Section A.2. We will focus on the ability of the Bayesian stochastic model specification search to detect time variation in the stochastic absorptive capacity parameters and on the small sample performance of inference on the unknown parameters in the most general model specification presented in Section 4.1 of the main paper.

Design

To make sure that our simulation results are relevant for putting the estimation results presented in the main paper in perspective, we simulate data for exactly the same sample size ($T = 62, N = 31$) that is available to us while the data generating process for log TFP $\lambda_{it} = y_{it} - \beta_i k_{it}$ is chosen to match with

the properties of the observed data. More specifically, we simulate data for λ_{it} from

$$\lambda_{it} = \alpha_{it} + \theta_{it} f_t + \varepsilon_{it}, \tag{A.22}$$

where

$$\alpha_{it} = \alpha_{i0} + \delta_\alpha \sqrt{\varsigma_\alpha} \widetilde{\alpha}_{it}, \qquad \text{with } \widetilde{\alpha}_{it} = \widetilde{\alpha}_{it-1} + \widetilde{\psi}^{\alpha}_{it}, \quad \widetilde{\alpha}_{i0} = 0, \quad \widetilde{\psi}^{\alpha}_{it} \sim \mathcal{N}(0,1), \tag{A.23}$$

$$\theta_{it} = \theta_{i0} + \delta_\theta \sqrt{\varsigma_\theta} \widetilde{\theta}_{it}, \qquad \text{with } \widetilde{\theta}_{it} = \widetilde{\theta}_{it-1} + \widetilde{\psi}^{\theta}_{it}, \quad \widetilde{\theta}_{i0} = 0, \quad \widetilde{\psi}^{\theta}_{it} \sim \mathcal{N}(0,1), \tag{A.24}$$

$$f_t = c + f_{t-1} + \sqrt{\varsigma_f} \psi^f_t, \qquad \psi^f_t \sim \mathcal{N}(0,1). \tag{A.25}$$

and $\alpha_{i0} \sim \mathcal{N}(0, \sigma^2_\alpha)$, $\theta_{i0} \sim \mathcal{N}(1, \sigma^2_\theta)$, $\varepsilon_{it} \sim \mathcal{N}(0, \sigma^2_\varepsilon)$.

Values for the unknown parameters are taken from the baseline estimation results presented in Section 3.3 of the main paper. First we set $\sigma_\varepsilon = 0.01$. Second, from the estimated values for α_{i0} and θ_{i0} we calculate their standard deviation over countries to be $\sigma_\alpha = 0.62$ and $\sigma_\theta = 0.24$. Third, we generate f_t from equation (A.25) using $c = 0.014$ and $\sqrt{\varsigma_f} = 0.013$. These values are obtained by fitting a random walk with drift to the cross-sectional average $\overline{\lambda}_t$ of λ_{it} (which is our proxy for f_t). Finally, we set $\sqrt{\varsigma_\alpha} = \sqrt{\varsigma_\theta} = 0.02$, which is in line with the unrestricted estimated standard deviation of the innovations to α_{it} reported in Table 2 of the main paper.

We consider four different scenarios concerning the presence of time variation in the absorptive capacity parameters α_{it} and θ_{it}, i.e., we set $(\delta_\alpha, \delta_\theta)$ equal to $(0,0)$, $(1,1)$, $(1,0)$ and $(0,1)$. This implies that α_{it} and θ_{it} are constant in the first scenario, while both are time varying in the second scenario. In the third and fourth scenario, there is only time variation in α_{it} and θ_{it} respectively. We generate 2,000 samples for each of the four scenarios. For each of the generated samples, we estimate the baseline and extended model using the MCMC algorithm outlined in Section A.2 with 50,000 draws of which the first 10,000 are discarded as burn-in.

Performance in the baseline model

We first estimate the baseline model outlined in Section 2 of the main paper to examine to what extent our Bayesian model selection approach is able to discriminate between the four different scenarios of time variation in α_{it} and θ_{it} and its performance for estimating $\sqrt{\varsigma_\alpha}$ and $\sqrt{\varsigma_\theta}$. Simulation results for each of the four scenarios are reported in Table A.1. We report selection fractions and average posterior inclusion probabilities for the binary indicators δ_α and δ_θ and for their combinations, together with the mean and standard deviation (over the Monte Carlo samples) of the median of the posterior distributions of $\sqrt{\varsigma_\alpha}$ and $\sqrt{\varsigma_\theta}$ (see notes to Table A.1 for details). The results show that our Bayesian model selection approach is very effective in detecting time variation. In each of the four scenarios, it picks the correct model in 96% or more of the samples. Moreover, the mean of the estimates for $\sqrt{\varsigma_\alpha}$ and $\sqrt{\varsigma_\theta}$ is always

close to the true value of 0 in case of no time variation and 0.02 in case of time variation.

Table A.1: Monte Carlo simulation results - Baseline model

Scenarios	Model selection frequencies						Posterior distribution	
	Indicators		Models $(\delta_\alpha, \delta_\theta)$				parameter	
$(\delta_\alpha, \delta_\theta)$	δ_α	δ_θ	(0,0)	(1,0)	(0,1)	(1,1)	$\sqrt{\varsigma_\alpha}$	$\sqrt{\varsigma_\theta}$
$(0,0)$	0.00	0.00	1.00	0.00	0.00	0.00	0.0000	0.0000
	[0.00]	[0.00]	[1.00]	[0.00]	[0.00]	[0.00]	(0.0000)	(0.0000)
$(1,0)$	1.00	0.01	0.00	0.99	0.00	0.01	0.0204	0.0001
	[1.00]	[0.03]	[0.00]	[0.97]	[0.00]	[0.03]	(0.0006)	(0.0009)
$(0,1)$	0.04	1.00	0.00	0.00	0.96	0.04	0.0005	0.0202
	[0.07]	[1.00]	[0.00]	[0.00]	[0.93]	[0.07]	(0.0022)	(0.0007)
$(1,1)$	0.98	1.00	0.00	0.00	0.02	0.98	0.0201	0.0196
	[0.98]	[1.00]	[0.00]	[0.00]	[0.02]	[0.98]	(0.0036)	(0.0026)

Notes: Based on 2,000 samples of size $N = 31$ and $T = 62$ generated using equations (A.22)-(A.25). For each sample, we estimate the baseline model presented in Section 2 of the main paper using MCMC with $50,000$ iterations after a burn-in of $5,000$ draws. The model selection frequencies are calculated as the fraction a certain indicator or model is selected (i.e., has a posterior inclusion probability larger than 0.5) over the Monte Carlo samples. In each sample, the posterior inclusion probabilities are calculated as the fraction of MCMC draws in which the stochastic model specification search prefers a model which allows for time variation in the corresponding parameter. Values in square brackets are average posterior inclusion probabilities over the Monte Carlo samples. For the parameters $\sqrt{\varsigma_\alpha}$ and $\sqrt{\varsigma_\theta}$, reported are the mean over the Monte Carlo samples of the median of the posterior distribution obtained using MCMC in each sample. Note that when an indicator δ_α or δ_θ is zero, the corresponding standard deviation $\sqrt{\varsigma_\alpha}$ or $\sqrt{\varsigma_\theta}$ is also set to zero. The values in brackets are the standard deviations of the median over the Monte Carlo samples.

Performance of the extended approach

Second, we use the data as simulated from equations (A.22)-(A.25) but now treat α_{it} and θ_{it} as observed data. In particular, we set $x_{1,it} = \theta_{it} - \frac{1}{N}\sum_{i=1}^{N}\theta_{it} - \frac{1}{T}\sum_{t=1}^{T}\theta_{it} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\theta_{it}$ and $x_{2,it} = \alpha_{it} - \frac{1}{N}\sum_{i=1}^{N}\alpha_{it} - \frac{1}{T}\sum_{t=1}^{T}\alpha_{it} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\alpha_{it}$ such that equation (A.22) can be rewritten as

$$\lambda_{it} = \alpha^*_{it} + \beta_{2\alpha}x_{2,it} + (\theta^*_{it} + \beta_{1\theta}x_{1,it})f_t + \varepsilon_{it}, \tag{A.26}$$

with $\alpha^*_{it} = \frac{1}{N}\sum_{i=1}^{N}\alpha_{it} + \frac{1}{T}\sum_{t=1}^{T}\alpha_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\alpha_{it}$ and $\theta^*_{it} = \frac{1}{N}\sum_{i=1}^{N}\theta_{it} + \frac{1}{T}\sum_{t=1}^{T}\theta_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\theta_{it}$ capturing the means removed from the simulated α_{it} and θ_{it} when constructing $x_{1,it}$ and $x_{2,it}$. The values for $\beta_{2\alpha}$ and $\beta_{1\theta}$ depend on whether there is time variation in α_{it} and θ_{it} respectively. If there is time variation in α_{it} (i.e., $\delta_\alpha = 1$), $\beta_{2\alpha} = 1$, while $\beta_{2\alpha} = 0$ otherwise. If there is time variation in θ_{it} (i.e., $\delta_\theta = 1$), $\beta_{1\theta} = 1$ while $\beta_{1\theta} = 0$ otherwise. Note that $\beta_{1\alpha}$ and $\beta_{2\theta}$ are always zero.

We estimate the extended specification presented in Section 4.1 of the main paper. For the current two-variable case, this is given by

$$\lambda_{it} = \alpha^*_{it} + \beta_{1t}x_{1,it} + (\beta^*_{2\alpha} + \beta_{1t}\beta^*_{2\theta})x_{2,it} + \theta^*_{it}\overline{y}_t + \varepsilon_{it}. \tag{A.27}$$

with $\beta^*_{2\theta} = \beta_{2\theta}/\beta_{1\theta} = 0$ and the four different scenarios for time variation in α_{it} and θ_{it} as outlined in the design above implying the following for $\beta_{1t} = (\overline{y}_t - \overline{a}^*_t)\beta_{1\theta}/\overline{\vartheta}^*_t$ and $\beta^*_{2\alpha} = \beta_{2\alpha}$:

1. $(\delta_\alpha, \delta_\theta) = (0,0)$: $\beta_{1t} = 0$ is constant and $\beta^*_{2\alpha} = 0$

2. $(\delta_\alpha, \delta_\theta) = (1, 0)$: $\beta_{1t} = 0$ is constant and $\beta^*_{2\alpha} = 1$

3. $(\delta_\alpha, \delta_\theta) = (0, 1)$: $\beta_{1t} = (\overline{y}_t - \overline{a}^*_t)/\overline{\vartheta}^*_t$ is time varying and $\beta^*_{2\alpha} = 0$

4. $(\delta_\alpha, \delta_\theta) = (1, 1)$: $\beta_{1t} = (\overline{y}_t - \overline{a}^*_t)/\overline{\vartheta}^*_t$ is time varying and $\beta^*_{2\alpha} = 1$

Simulation results for estimating the unrestricted model in equation (A.27) under each of the four scenarios of time variation are reported in the left panel of Table A.2. In particular, we report the fraction of samples in which β_{1t} is chosen to be time varying by the Bayesian model selection algorithm together with the mean and standard deviation (over the Monte Carlo samples) of the median of the posterior distributions of $\beta^*_{2\alpha}$ and $\beta^*_{2\theta}$ (see notes to Table A.2 for details). The results show that our Bayesian model selection approach is very effective in detecting time variation in β_{1t} (and hence testing whether $\beta_{1\theta}$ is zero or not). In each of the four scenarios, it always picks the correct model. Moreover, the mean of the estimates for $\beta^*_{2\alpha}$ and $\beta^*_{2\theta}$ is always close to the true value of 0 for $\beta^*_{2\theta}$ and 0 (when $\delta_\alpha = 0$) or 1 (when $\delta_\alpha = 1$) for $\beta^*_{2\alpha}$. In the right panel of Table A.2, we report estimates of the restricted version of the extended model presented in equation (24) of Section 4.1. This model is only valid in the first two scenarios where $\delta_\theta = 0$ as it requires $\beta_{k\theta} = 0$ $(\forall k)$. The results show that the mean of the estimates for $\beta_{1\alpha}$ and $\beta_{2\alpha}$, which are now identified, is always close to the true value of 0 for $\beta_{1\alpha}$ and 0 (when $\delta_\alpha = 0$) or 1 (when $\delta_\alpha = 1$) for $\beta_{2\alpha}$.

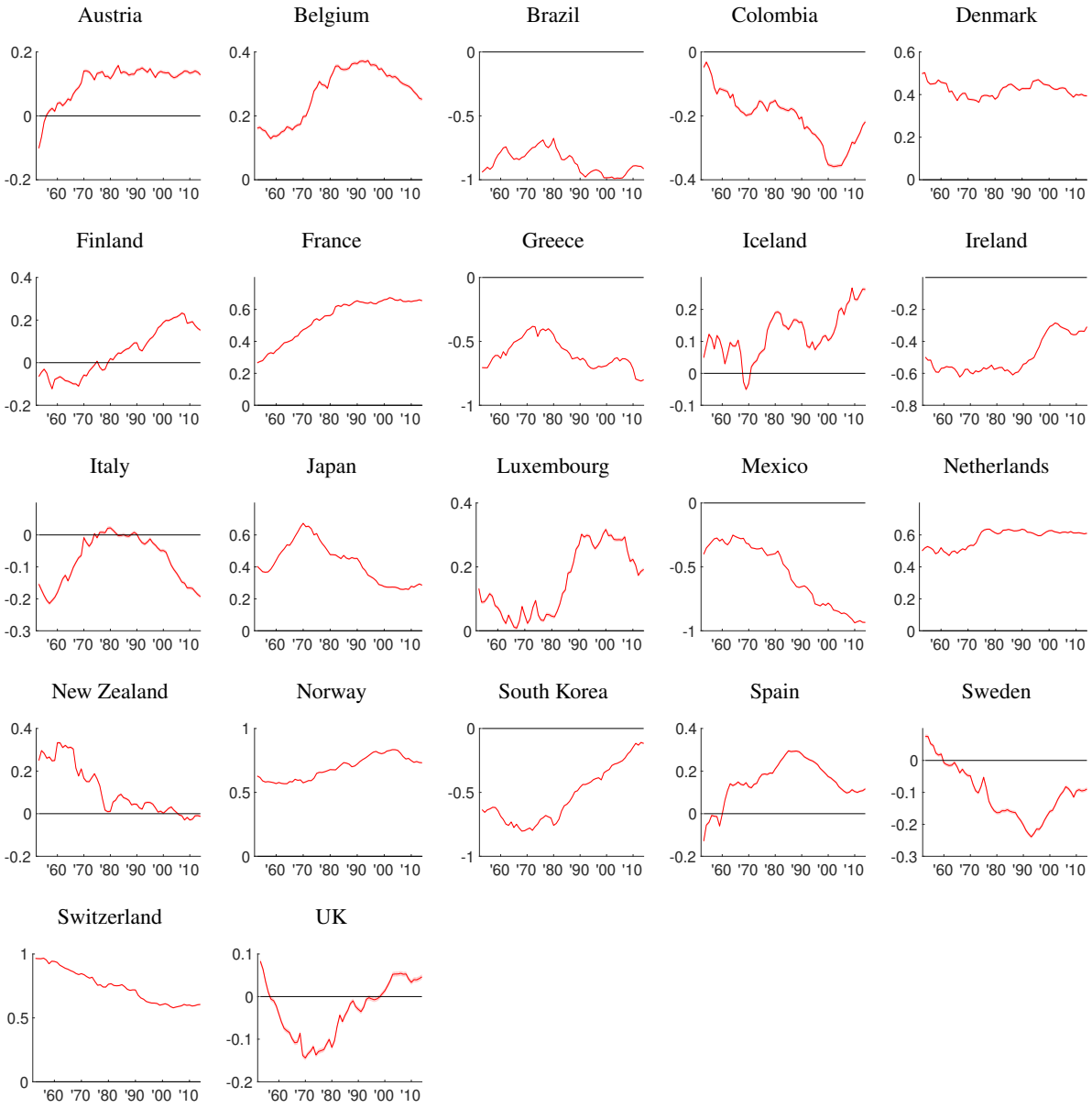
Table A.2: Monte Carlo simulation results - Extended model

	Unrestricted model			Restricted model	
Scenarios	Indicator	Posterior distribution parameters			
$(\delta_\alpha, \delta_\theta)$	δ_β	$\beta^*_{2\alpha}$	$\beta^*_{2\theta}$	$\beta_{1\alpha}$	$\beta_{2\alpha}$
$(0, 0)$	0.00	0.00	0.00	0.00	0.00
	[0.01]	(0.01)	(0.00)	(0.01)	(0.01)
$(1, 0)$	0.00	1.00	0.00	0.00	1.00
	[0.01]	(0.01)	(0.00))	(0.01)	(0.01)
$(0, 1)$	1.00	0.00	0.00		
	[1.00]	(0.03)	(0.02)		
$(1, 1)$	1.00	1.00	0.00		
	[1.00]	(0.03)	(0.03)		

Notes: Based on 2,000 samples of size $N = 31$ and $T = 62$ generated using equations (A.22)-(A.25). For each sample, we estimate the extended model presented in Section 4.1 of the main paper using MCMC with $50,000$ iterations after a burn-in of $5,000$ draws. For the indicator δ_β we report the frequency β_{1t} is chosen to be time varying (i.e. has a posterior inclusion probability larger than 0.5) over the Monte Carlo samples. In each sample, the posterior inclusion probability is calculated as the fraction of MCMC draws in which a model with time variation in β_{1t} is preferred. Values in square brackets are average posterior inclusion probabilities over the Monte Carlo samples. The summary statistics for the parameters are the mean over the Monte Carlo samples of the median of the posterior distribution obtained using MCMC in each sample. The values in brackets are the standard deviations of the median over the Monte Carlo samples.

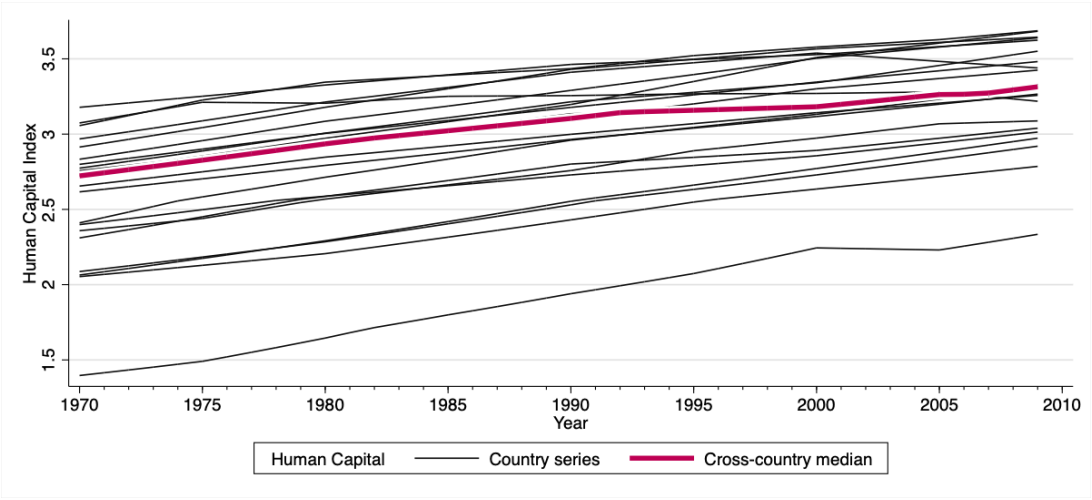
Appendix B Additional Figures

Figure B.1: Posterior results for the absorptive capacity parameter α_{it} ($\theta_{it} = 1$)



Notes: Reported are the posterior mean and 68% highest density interval (HDI) for a parsimonious-hybrid model setting $\theta_{it} = \theta_i$ and additionally $\theta_i = 1$ for those countries where the posterior probability that $\theta_i \neq 1$ is smaller than 0.5. Based on MCMC with $K = 200,000$ iterations where the first $B = 20,000$ are discarded as burn-in. The average integrated autocorrelation time across the plotted α's is 1.

Figure B.2: Human Capital Index (PWT)



Appendix C Case Studies of Structural and Economic Reforms

C.1 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 in the main paper shows Ireland's absorptive capacity to be stable until the early 1970s. This, however, was not a favorable position 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 expansionary 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 targeting successfully attracted US high-tech companies searching for a European base. The resulting stream of incoming FDI fostered Ireland's stock of knowledge, which in turn led to a steep increase of its absorptive capacity.

C.2 Sweden

Sweden's absorptive capacity evolution is characterized by a moderate deterioration 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 centralized negotiations came into play and a higher degree of regulation applied to capital and labor 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 breakthrough of microelectronics. Instead of transforming the economy throughout the 1970s and 1980s, the focus of successive governments 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 government 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.

C.3 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 limited number of industries in which it believed to possess a comparative advantage rather than allow for a market-based orientation. To this end the government 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 manufacturing technology in Europe and North America, fueled 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. Information and Communication Technologies (ICT) only gradually found their way to Japanese firms as high job security made it difficult for companies to shed unskilled labor Moreover, Japan is facing an aging working force further holding down productivity growth.

References

Bitto, A. and Frühwirth-Schnatter, S. (2019). Achieving shrinkage in a time-varying parameter model framework. *Jounal of Econometrics*, 210:75–97.

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

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.

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.

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.