An Econometric Model of International Long-run Growth Dynamics

Supplementary Material

Ulrich K. Müller
Princeton University

James H. Stock
Harvard Unversity

Mark W. Watson
Princeton University

This Draft: October 10, 2020

Appendix 1 Data

The data are annual values of real per-capita GDP for 113 countries spanning the 118-year period 1900-2017, taken from the Penn Word Table 9.1 (Feenstra, et al. (2015) and Inklar and Woltjer (2019)) and Bolt et al. (2018) Maddison Project Database. GDP is measured at constant 2011 national prices, expressed in U.S. dollars.

From the Penn World Table, we used the real GDP series $rgdp^{na}$, which measures real GDP at constant national prices, and is obtained from national accounts data for each country. We also extracted the population series *population* from the PWT. From these, we computed $rgdpnapc = rgdp^{na}/population$ for each year and country.

From the Maddison Project Database we extracted *rgpdnapc*, which is the same concept as $rgpd^{na}$ in PWT, but expressed in per-capita terms. We also extracted the population series *pop*.

We then linked the series for each country as follows as follows: Letting i denote country, t denote year, and t_i denote the first year that *rgdpnapc* is available in the PWT. Then

$$rgdpnapc_{i,t}^{linked} = \begin{cases} rgdpnapc_{i,t}^{PWT} \text{ for } t \geq t_i \\ grpdnapc_{i,t}^{Maddison} \times (rgdpnapc_{i,t_i}^{PWT}/rgdpnapc_{i,t_i}^{Maddison}) \text{ for } t < t_i \end{cases}$$

Population was linked analogously. Results in the paper are based on these linked series.

The file **TableA1_List_of_Countries.xlsx** lists the 113 countries used in the analysis, each country's 2017-value of per-capita real GDP and population, and start and end dates. The file **TableA2_Excluded_Countries. xlsx** lists the 69 countries in the PWT that were excluded because of a short sample (less than 50 years), a small 2017 population (less than 3 million) or both.

The sample ends in 2017 (the last year available in PWT 9.1) and begins in 1900, where the (arbitrary) start date means that the sample contains at most 118 years of data. While data are available for some countries in earlier years from the Maddison Project Database, concerns about temporal stability led us to start the sample in 1900.

The results discussed in Section 6.3 use the same measure of real GDP from the Penn World Table, but replace the population series with employment (*emp*).

Appendix 2 Computations

Appendix 2.1 Unbalanced Panel

Müller and Watson (2008, 2018) extract low-frequency trends by projecting the original time series y_t onto a finite set of trigonometric low-frequency regressors. In particular, for time series that have a deterministic trend, their suggestion for the regressors are the (asymptotic version) of the eigenvectors associated with the largest eigenvalues of a demeaned and detrended random walk. The variation in a random walk is dominated by low-frequency components (cf. Phillips (1998)). Thus, high variance linear combinations of a (demeaned and detrended) random walk extract low frequency variability, and the magnitude of the variance becomes an indicator for the associated frequency.

To be specific, let V be the $T \times T$ covariance matrix of a random walk with unit variance and initialized at $t = 0$, that is, $V_{ij} = \min(i, j)$; let R denote the $T \times 2$ matrix with tth row equal to $(1, t - (T + 1)/2)$. The full sample regressors that extract variation below the frequency cut-off $2q/T$ are then given by $q + 1$ vectors: Two corresponding to the columns of R, plus the $q - 1$ eigenvectors of the $T \times T$ matrix

$$MVM \quad \text{with } M = I_T - R(R'R)^{-1}R' \tag{1}$$

corresponding to the $q - 1$ largest eigenvalues. As discussed in Müller and Watson (2008), the cut-off formula $2q/T$ requires counting the the linear trend in the second column of R as part of the q non-constant vectors that extract variation. For future reference, denote the $q - 1$ largest eigenvalue of MVM by κ_q. Note that the regressors are orthogonal to each other by construction.

We now generalize this approach further and make it operational in an unbalanced panel. Let J_i be a diagonal $T \times T$ matrix whose tth diagonal element is equal to one if the data $y_{i,t}$ for country i is observed at time t, and zero otherwise. Let $R_i = J_i R$, a $T \times 2$ matrix with zeros in all rows where $y_{i,t}$ is unobserved. The regressors for series $y_{i,t}$ are then given by $q_i + 1$ weights: The two columns of R_i, and the $q_i - 1$ eigenvectors of

$$M_i V M_i \;\; \text{with } M_i = J_i - R_i(R_i'R_i)^{-1}R_i' \tag{2}$$

with eigenvalues larger or equal than κ_q. Note that the eigenvectors of $M_i V M_i$ are equal to zero by construction in all rows where $y_{i,t}$ is unobserved. Also note that if $y_{i,t}$ is observed for a single spell of T_i consecutive periods, then this construction yields the same set of regressors that would have been obtained from (1) for a fully observed time series of length T_i with cut-off frequency $2q_i/T_i \approx 2q/T$.

Appendix 2.2 Low-Frequency Trends and their Asymptotic Distribution

Let the $q_i + 1$ vector Y_i be the OLS coefficients of a regression of $\{y_{i,t}\}_{t=1}^T$ on the $q_i + 1$ regressors obtained in this manner. The low-frequency panel analysis treats Y_i as the only observation from country i, ignoring all higher frequency sample variation contained in the original series $\{y_{i,t}\}_{t=1}^T$.

The unbalanced panel introduces a complication in preserving the factor structure (1) for the low frequency trends for $y_{i,t}$ and the common factor f_t. We procede by introducing a common higher dimensional "baseline" low-frequency trend X_i, of which Y_i is a partial observation (even if there is no missing data for country i). Specifically, consider the space spanned by the $\tilde{q}+1$ regressors extracted from (1), with $\tilde{q} >> q$. For sufficiently large $\tilde{q}$, we can approximate the q_i+1 regressors obtained from (2) arbitrarily well, that is, the regressors obtained from (2) can be written as a linear combination of the $\tilde{q} + 1$ baseline regressors. For computational efficiency, it is not attractive to make $\tilde{q}$ too large; we choose $\tilde{q} = q + 15$ throughout. We then use ordinary least squares regressions (with T observations) to find the best linear combination to approximate the $q_i + 1$ regressors as linear combinations of the $\tilde{q}+1$ baseline regressors (and to make this approximation more accurate, we treat as missing any isolated observed value, that is a value $y_{i,t}$ where both $y_{i,t-1}$ and $y_{i,t+1}$ are missing). The smallest R^2 across all regressors and countries in this construction is larger than 0.973.

With this approximation in place, the relationship between the observed $q_i + 1$ vector Y_i and the latent augmented $\tilde{q} + 1$ trend X_i is given by

$$Y_i = B_i' X_i \tag{3}$$

where B_i is the $\tilde{q} + 1 \times q_i + 1$ matrix obtained from the OLS regressions just described.

The linear structure of the model of equations (1)-(7) then yields a corresponding linear structure in the $\tilde{q} + 1$ dimensional low-frequency trends

$$X_i \quad = \quad F + C_i, \;\; i = 1, \ldots, n \tag{4}$$

$$\begin{aligned}
C_i &= \mu_c \iota_1 + \lambda_{c,i} G_{J(i)} + U_{c,i}, \quad i = 1, \ldots, n & (5) \\
G_j &= \lambda_{g,j} H_{K(j)} + U_{g,j}, \;\; j = 1, \ldots, 25 & (6) \\
H_k &= U_{h,k}, \quad k = 1, \ldots, 10 & (7) \\
F &= \iota_1 f_0 + \iota_2 \mu_m + S_m + A & (8) \\
Y^0 &= \sum_{i=1}^{n} w_i X_i - F & (9)
\end{aligned}$$

where ι_1 and ι_2 in (5) and (8) are the first and second column of $I_{\tilde{q}+1}$, respectively, and F, C_i, $U_{c,i}$, $U_{g,j}$, $U_{h,k}$, G_j, H_k, S_m and A are the $\tilde{q}+1$ low-frequency trends obtained from f_t, $c_{i,t}$, $u_{c,i,t}$, $u_{g,j,t}$, $u_{h,k,t}$, $g_{j,t}$, $h_{k,t}$, m_t and Δa_t, respectively. Equation (9) here represents the in-sample prior that f_t is close to the population weighted average of OECD countries' $y_{i,t}$, implemented via a dummy observation prior where $Y^0 \sim \mathcal{N}(0, \Delta)$ with $\Delta = 0.01^2 I_{\tilde{q}+1}$ in the prior, and the observed realization is $Y^0 = 0$. The population weights w_i in (9) sum to one and are set to zero for non-OECD countries.

Now as discussed in Section 4.2, the focus on low-frequency variation offers the advantage that the large sample distribution of the low-frequency trends $U_{c,i}$, $U_{G,j}$, $U_{h,k}$, S_m and A of the driving innovations of the model are approximately Gaussian, with the large sample covariance matrix a function of the low-frequency scale and persistence parameters. For instance, recall from (8) that the u_t terms in (2)-(4) are modelled as a weighted sum of two independent low-frequency (=local-to-unity) stationary AR(1) processes with unit long-run variance. Under such asymptotics, the low-frequency trend U computed from $\{u_t\}_{t=1}^T$ satisfies $U \overset{a}{\sim} \mathcal{N}(0, \sigma_u^2 \Sigma_U(\theta^u))$ with $\theta^u = (\rho_1, \rho_2, \zeta)$, that is, the limiting Gaussian distribution only depends on low-frequency parameters and not, say, on the short-run dynamics of the disturbances driving the local-to-unity processes. In other words, under such asymptotics, U is distributed *as if* the underlying u_t was a sum of two independent Gaussian AR(1) processes with coefficient ρ_1 and ρ_2, respectively. Thus, an asymptotically justified approximation to the the $\tilde{q}+1 \times \tilde{q}+1$ covariance matrix $\Sigma_U(\theta^u)$ is simply given by

$$\Sigma_U(\theta^u) = (\tilde{R}'\tilde{R})^{-1} \tilde{R}' \tilde{\Sigma}_u(\theta^u) \tilde{R} (\tilde{R}'\tilde{R})^{-1}$$

where $\tilde{R}$ is the $T \times (\tilde{q}+1)$ dimensional matrix of baseline regressors, and $\tilde{\Sigma}^u(\theta_u)$ is the $T \times T$ covariance matrix of a $(\zeta, \sqrt{1-\zeta^2})$ weighted average of two independent stationary AR(1) processes with coefficients ρ_1 and ρ_2 and innovation variance equal to $(1-\rho_1^2)^{-1/2}$

and $(1-\rho_2^2)^{-1/2}$, respectively. The same arguments also yield

$$S_m \overset{a}{\sim} \mathcal{N}(0, \sigma_m^2 \Sigma_m(\rho_m)) \tag{10}$$

$$A \overset{a}{\sim} \mathcal{N}(0, \sigma_{\Delta a}^2 \Sigma_a) \tag{11}$$

where

$$\begin{aligned} \Sigma_m(\rho_m) &= (\tilde{R}'\tilde{R})^{-1}\tilde{R}'\tilde{\Sigma}_m(\rho_m)\tilde{R}(\tilde{R}'\tilde{R})^{-1} \\ \Sigma_a &= (\tilde{R}'\tilde{R})^{-1}\tilde{R}'V\tilde{R}(\tilde{R}'\tilde{R})^{-1} \end{aligned}$$

with $\tilde{\Sigma}_m(\rho_m)$ the $T \times T$ covariance matrix of the partial sum of a stationary Gaussian mean-zero AR(1) model with coefficient ρ_m and innovation variance equal to $(1-\rho_m^2)^{-1}$ (so that the unconditional variance of the AR(1) process is normalized to be equal to unity), and V is the random-walk covariance matrix introduced earlier.

Further note that the joint large-sample distribution of the in-sample low-frequency trends and out-of-sample long-horizon forecasts is again normal, as discussed and exploited in Müller and Watson (2016). Taking the example of the forecast of $u_{T+\lfloor hT \rfloor}$ for some sample size independent $h > 0$, the joint $(\tilde{q}+2) \times (\tilde{q}+2)$ approximate covariance matrix of $(\tilde{R}'\tilde{R})^{-1}\tilde{R}'u_{1:T}$ and $u_{T+\lfloor hT \rfloor}$ may be computed via

$$\tilde{W}^{e\prime}\tilde{\Sigma}_u^e(\theta^u)\tilde{W}^e \text{ with } \tilde{W}^{e\prime} = \begin{pmatrix} (\tilde{R}'\tilde{R})^{-1}\tilde{R}' & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \end{pmatrix} \tag{12}$$

where $\tilde{W}^e$ is $(T+\lfloor hT \rfloor) \times (\tilde{q}+2)$ and $\tilde{\Sigma}_u^e(\theta^u)$ is the $(T+\lfloor hT \rfloor) \times (T+\lfloor hT \rfloor)$ covariance matrix of a $(\zeta, \sqrt{1-\zeta^2})$ weighted average of two independent stationary AR(1) processes with coefficients ρ_1 and ρ_2 and innovation variance equal to $(1-\rho_1^2)^{-1/2}$ and $(1-\rho_2^2)^{-1/2}$, respectively. The same approximation readily extends to the forecast of $f_{T+\lfloor hT \rfloor} = \sum_{t=1}^{T+\lfloor hT \rfloor}(m_s + \Delta a_s)$.

Appendix 2.3 Gibbs Sampler

The posterior results in the paper are computed from a (thinned) version of 150,000 Gibbs draws, after discarding 25,000 for burn-in. Given our choice of prior, each step of the Gibbs sampler is conjugate. Further, the dimension reduction to $\tilde{q}+1$ latent vectors, and the discrete parameter space allows precomputations of nearly all relevant $(\tilde{q}+1) \times (\tilde{q}+1)$ matrices. Finally, many of the computations within one Gibbs draw can be parallelized, as they involve n, 25 or at least 10 independent components. In combination, this leads to very fast computations: In the baseline specification, the 175,000 draws take about 4 minutes on a dual 12 core workstation in a Fortran implementation.

The tight prior variance Δ of the dummy observation Y^0 in (9) prevents large moves of F; in order to accelerate convergence to reasonable starting values, we linearly decrease the variance of Y^0 from 1000Δ to Δ during the burn-in phase.

To complete the description of the model, recall that the innovations $\{U_{c,i}\}_{i=1}^n$, $\{U_{g,j}\}_{j=1}^{25}$ and $\{U_{h,k}\}_{k=1}^{10}$ are independent with distribution

$$\begin{aligned} U_{c,i} &\sim \mathcal{N}(0,\omega^2\kappa_{c,i}^2(1-\lambda_{c,i}^2)\Sigma_U(\theta_{c,i}^u)) &(13)\\ U_{g,j} &\sim \mathcal{N}(0,\omega^2\kappa_{g,j}^2(1-\lambda_{g,j}^2)\Sigma_U(\theta_{g,j}^u)) &(14)\\ U_{h,k} &\sim \mathcal{N}(0,\omega^2\kappa_{h,k}^2\Sigma_U(\theta_{h,k}^u)) &(15)\end{aligned}$$

where

- $\{\lambda_{c,i}\}_{i=1}^n$ and $\{\lambda_{g,j}\}_{j=1}^{25}$ are i.i.d. draws from the discrete distribution with 25 support points $\{\lambda^l\}_{l=1}^{25}$ and priors $p_c^\lambda=(p_{c,1}^\lambda,\ldots,p_{c,25}^\lambda)$ and $p_g^\lambda=(p_{g,1}^\lambda,\ldots,p_{g,25}^\lambda)$, respectively
- $\{\theta_{c,i}^u\}_{i=1}^n$, $\{\theta_{g,j}^u\}_{j=1}^{25}$ and $\{\theta_{h,k}^u\}_{k=1}^{10}$ are i.i.d. draws from the discrete distribution with 100 support points $\{\theta^l\}_{l=1}^{100}$ and priors $p_c^u=(p_{c,1}^u,\ldots,p_{c,100}^u)$, $p_g^u=(p_{g,1}^u,\ldots,p_{g,100}^u)$ and $p_h^u=(p_{h,1}^u,\ldots,p_{h,100}^u)$, respectively
- $\{\kappa_{c,i}\}_{i=1}^n$, $\{\kappa_{g,j}\}_{j=1}^{25}$, and $\{\kappa_{h,k}\}_{k=1}^{10}$ are i.i.d. draws from the discrete distribution with 25 support points $\{\kappa^l\}_{l=1}^{25}$ and priors $p_c^\kappa=(p_{c,1}^\kappa,\ldots,p_{c,25}^\kappa)$, $p_g^\kappa=(p_{g,1}^\kappa,\ldots,p_{g,25}^\kappa)$ and $p_h^\kappa=(p_{h,1}^\kappa,\ldots,p_{h,25}^\kappa)$, respectively.

and the various p are realizations of the corresponding Dirichlet prior.

The state of the sampler thus consists of $\{X_i\}_{i=1}^n\in\mathbb{R}^{n(\tilde q+1)}$, $(F,S_m)\in\mathbb{R}^{2(\tilde q+1)}$, $\{G_j\}_{j=1}^{25}\in\mathbb{R}^{25(\tilde q+1)}$, $\{H_k\}_{k=1}^{10}\in\mathbb{R}^{10(\tilde q+1)}$, $\mu_c\in\mathbb{R}$, $\{\lambda_{c,i}\}_{i=1}^n\in\mathbb{R}^n$, $\{\lambda_{g,j}\}_{j=1}^{25}\in\mathbb{R}^{25}$, $\{\theta_{c,i}^u\}_{i=1}^n\in\mathbb{R}^{3n}$, $\{\theta_{g,j}^u\}_{j=1}^{25}\in\mathbb{R}^{75}$, $\{\theta_{h,k}^u\}_{k=1}^{10}\in\mathbb{R}^{30}$, $\omega^2\in\mathbb{R}$, $\{\kappa_{c,i}\}_{i=1}^n\in\mathbb{R}^n$, $\{\kappa_{g,j}\}_{j=1}^{25}\in\mathbb{R}^{25}$, $\{\kappa_{h,k}\}_{k=1}^{10}\in\mathbb{R}^{10}$, $\{J(i)\}_{i=1}^n\in\mathbb{N}^n$, $\{K(j)\}_{i=1}^{25}\in\mathbb{N}^{25}$, $(f_0,\mu_m)\in\mathbb{R}^2$, $\sigma_{\Delta a}^2\in\mathbb{R}$, $\sigma_m\in\mathbb{R}$, $\rho_m\in\mathbb{R}$, $p_c^\lambda\in\mathbb{S}^{25}$, $p_g^\lambda\in\mathbb{S}^{25}$, $p_c^\theta\in\mathbb{S}^{100}$, $p_g^\theta\in\mathbb{S}^{100}$, $p_h^\theta\in\mathbb{S}^{100}$, $p_c^\kappa\in\mathbb{S}^{25}$, $p_g^\kappa\in\mathbb{S}^{25}$, $p_h^\kappa\in\mathbb{S}^{25}$, where $\mathbb{S}^m$ is the simplex of dimension m. In each step of the Gibbs sampler described in detail below, one of these states is drawn conditional on the current value of all other states. We checked the correctness of the sampler (and code) using the Geweke (2004) test as described in Müller and Watson (2019), who also provide additional details on the conjugate distributions exploited in the Gibbs steps below. The long-run forecasts are easily generated conditional on the full state of the sampler from the conditional normal distribution with parameters implied by (12) and the analogous expression for the forecast of f_t.

One draw of the Gibbs sampler consists of the following 28 steps.

1. $\{X_i\}_{i=1}^n$

   From (4), (5) and (9), drawing $\{X_i\}_{i=1}^n$ amounts to drawing from the conditional distribution of $\{C_i\}_{i=1}^n$ given observations $\{B_i'C_i\}_{i=1}^n$ and $Y^0 = \sum_{i=1}^n w_i C_i$, where $C_i = X_i - F$. Let $\mu_C = (\mu_c \iota_1' + \lambda_{c,1} G'_{J(1)}, \ldots, \mu_c \iota_1' + \lambda_{c,n} G'_{J(n)})' \in \mathbb{R}^{n(\tilde{q}+1)}$ be the conditional mean of $C = (C_1', \ldots, C_n')'$, Σ the block diagonal conditional covariance matrix of U_c with ith block equal to $\omega^2 \kappa_{c,i}^2 (1-\lambda_{c,i}^2) \Sigma_U(\theta_{c,i}^u)$, B the $n(\tilde{q}+1) \times \sum_{i=1}^n (q_i+1)$ matrix such that $Y = (Y_1', \ldots, Y_n') = B'X$ with $X = (X_1', \ldots, X_n')'$, and $w = (w_1, \ldots, w_n)' \otimes I_{\tilde{q}+1} \in \mathbb{R}^{n(\tilde{q}+1) \times (\tilde{q}+1)}$. The conditional joint distribution of C, Y_0 and $B'C$ is then given by

$$\begin{pmatrix} C \\ Y^0 \\ B'C \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_C \\ w'\mu_C \\ B'\mu_C \end{pmatrix}, \begin{pmatrix} \Sigma & \cdot & \cdot \\ w'\Sigma & w'\Sigma w + \Delta & \cdot \\ B'\Sigma & B'\Sigma w & B'\Sigma B \end{pmatrix} \right)$$

   so that with $m = \mu_C + \Sigma B (B'\Sigma B)^{-1} B'(C - \mu_C)$ and $V = \Sigma - \Sigma B (B'\Sigma B)^{-1} B'\Sigma$,

$$\begin{pmatrix} C \\ Y^0 \end{pmatrix} | B'C \sim \mathcal{N}\left(\begin{pmatrix} m \\ w'm \end{pmatrix}, \begin{pmatrix} V & \cdot \\ w'V & w'Vw + \Delta \end{pmatrix} \right).$$

   Thus,

$$C|(B'C, Y^0) \sim \mathcal{N}(m + Vw(w'Vw+\Delta)^{-1}(Y^0 - w'm), V - V'w(w'Vw+\Delta)^{-1}w'V). \quad (16)$$

   In order to avoid manipulating the large matrices in this expression directly, we generate a draw from this distribution as follows: Draw $Z_0 \sim \mathcal{N}(\mu_C, \Sigma)$, and note that

$$Z_1 = Z_0 - \Sigma B (B'\Sigma B)^{-1} B'(Z_0 - C) \sim \mathcal{N}(m, V).$$

   The draw of Z_0 and the computation of Z_1 can be implemented independently for each $\tilde{q}+1$ block, and each $\tilde{q}+1 \times \tilde{q}+1$ block of $B(B'\Sigma B)^{-1}B'$ can be precomputed (up to scale). Then draw $\varepsilon \sim \mathcal{N}(Y^0, \Delta)$ independent of Z_0, and note that

$$Z_1 - Vw(w'Vw+\Delta)^{-1}w'(Z_1 - \varepsilon) \sim \mathcal{N}(m + Vw(w'Vw+\Delta)^{-1}(Y^0 - w'm), V - V'w(w'Vw+\Delta)^{-1}w'V)$$

which is a draw from the target (16), and the evaluation of Vw and $w'Vw$ simply consists of computing weighted sums of precomputed $\tilde{q}+1 \times \tilde{q}+1$ blocks.

2. (F, S_m)

   Let $\mu_F = \iota_1 f_0 + \iota_2 \mu_m$, $\Sigma_F = \sigma_m^2 \Sigma_m(\rho_m) + \sigma_{\Delta a}^2 \Sigma_a$ and $e = (1, \ldots, 1)' \otimes I_{\tilde{q}+1} \in \mathbb{R}^{n(\tilde{q}+1)\times(\tilde{q}+1)}$. Then from the independent Gaussian kernels (4) and (9), we obtain that the conditional posterior distribution of F is

   $$F - \mu_F \sim \mathcal{N}(m_F, V_F)$$

   where $V_F = (\Sigma_F^{-1} + e'\Sigma^{-1}e + \Delta^{-1})^{-1}$ and $m_F = V_F(e'\Sigma^{-1}(X - \mu_C - e\mu_F) + \Delta^{-1}(w'X - Y^0 - \mu_F))$. Furthermore, from (8), conditional on a draw of F from this conditional distribution, the conditional distribution of S_m satisfies

   $$S_m|F \sim \mathcal{N}(\Sigma_S \Sigma_F^{-1}(F - \mu_F), \Sigma_S - \Sigma_S \Sigma_F^{-1} \Sigma_S)$$

   where $\Sigma_S = \sigma_m^2 \Sigma_m(\rho_m)$.

3. $\{G_j\}_{j=1}^{25}$

   Let $\Sigma_{g,j} = \omega^2 \kappa_{g,j}^2 (1 - \lambda_{g,j}^2) \Sigma_U(\theta_{g,j}^u)$ and $\Sigma_{c,i} = \omega^2 \kappa_{c,i}^2 (1 - \lambda_{c,i}^2) \Sigma_U(\theta_{c,i}^u)$. From (5) and (6), we have that the conditional distribution of $\{G_j\}_{j=1}^{25}$ is independent and satisfies

   $$G_j \sim \mathcal{N}(m_{G_j}, V_{G_j})$$

   where $V_{G_j} = (\Sigma_{g,j}^{-1} + \sum_{i=1}^n \mathbf{1}[J(i) = j] \lambda_{c,i}^2 \Sigma_{c,i}^{-1})^{-1}$ and $m_{G_j} = V_{G_j}(\lambda_{g,j} \Sigma_{g,j}^{-1} H_{K(j)} + \sum_{i=1}^n \mathbf{1}[J(i) = j] \lambda_{c,i} \Sigma_{c,i}^{-1}(C_i - \mu_c \iota_1))$ with $C_i = X_i - F$.

4. $\{H_k\}_{k=1}^{10}$

   Let $\Sigma_{h,k} = \omega^2 \kappa_{h,k}^2 \Sigma_U(\theta_{h,k}^u)$ and $\Sigma_{g,j} = \omega^2 \kappa_{g,j}^2 (1 - \lambda_{g,j}^2) \Sigma_U(\theta_{g,j}^u)$. From (6) and (7), we have that the conditional distribution of $\{H_k\}_{j=1}^{25}$ is independent and satisfies

   $$H_k \sim \mathcal{N}(m_{H_k}, V_{H_k})$$

   where $V_{H_k} = (\Sigma_{h,k}^{-1} + \sum_{j=1}^{25} \mathbf{1}[K(j) = k] \lambda_{g,j}^2 \Sigma_{g,j}^{-1})^{-1}$ and $m_{H_k} = V_{H_k}(\sum_{j=1}^{25} \mathbf{1}[K(j) = k] \lambda_{g,j} \Sigma_{g,j}^{-1} G_j)$.

5. μ_c

Under a flat prior for μ_c, the conditional posterior distribution is

$$\mu_c \sim \mathcal{N}\left(\frac{\sum_{i=1}^n \iota_1' \Sigma_{c,i}^{-1}(C_i - \lambda_{c,i} G_{J(i)})}{\sum_{i=1}^n \iota_1' \Sigma_{c,i}^{-1} \iota_1}, \frac{1}{\sum_{i=1}^n \iota_1' \Sigma_{c,i}^{-1} \iota_1}\right)$$

where $\Sigma_{c,i} = \omega^2 \kappa_{c,i}^2 (1 - \lambda_{c,i}^2) \Sigma_U(\theta_{c,i}^u)$, and $C_i = X_i - F$.

6. $\{\lambda_{c,i}\}_{i=1}^n$

From (5) and (13), the conditional posterior weight of $\lambda_{c,i}$ on λ^l, $l = 1, \ldots, 25$ is proportional to

$$p_{c,l}^\lambda \exp[-\tfrac{1}{2}(C_i - \mu_c \iota_1 - \lambda^l G_{J(i)})'(\omega^2 \kappa_{c,i}^2 (1-(\lambda^l)^2) \Sigma_U(\theta_{c,i}^u))^{-1}(C_i - \mu_c \iota_1 - \lambda^l G_{J(i)})](1-(\lambda^l)^2)^{-(\tilde{q}+1)/2}$$

with $C_i = X_i - F$.

7. $\{\lambda_{g,j}\}_{j=1}^{25}$

From (6) and (14), the conditional posterior weight of $\lambda_{g,j}$ on λ^l, $l = 1, \ldots, 25$ is proportional to

$$p_{g,l}^\lambda \exp[-\tfrac{1}{2}(G_j - \lambda^l H_{K(j)})'(\omega^2 \kappa_{g,j}^2 (1-(\lambda^l)^2) \Sigma_U(\theta_{g,j}^u))^{-1}(G_j - \lambda^l H_{K(j)})](1-(\lambda^l)^2)^{-(\tilde{q}+1)/2}.$$

8. $\{\theta_{c,i}^u\}_{i=1}^n$

From (13), the conditional posterior weight of $\theta_{c,i}^u$ on θ^l, $l = 1, \ldots, 100$ is proportional to

$$p_{c,l}^\theta |\det \Sigma_U(\theta^l)|^{-1/2} \exp\left[-\frac{1}{2} \frac{U_{c,i}' \Sigma_U(\theta^l)^{-1} U_{c,i}}{\omega^2 \kappa_{c,i}^2 (1 - \lambda_{c,i}^2)}\right]$$

with $U_{c,i} = X_i - F - \mu_c \iota_1 - \lambda_{c,i} G_{J(i)}$.

9. $\{\theta_{g,j}^u\}_{j=1}^{25}$

From (14), the conditional posterior weight of $\theta_{g,j}^u$ on θ^l, $l = 1, \ldots, 100$ is proportional to

$$p_{g,l}^\theta |\det \Sigma_U(\theta^l)|^{-1/2} \exp\left[-\frac{1}{2} \frac{U_{g,j}' \Sigma_U(\theta^l)^{-1} U_{g,j}}{\omega^2 \kappa_{g,j}^2 (1 - \lambda_{g,j}^2)}\right]$$

with $U_{g,j} = G_j - \lambda_{g,j} H_{K(j)}$.

10. $\{\theta^u_{h,k}\}_{k=1}^{10}$

From (7) and (15), the conditional posterior weight of $\theta^u_{h,k}$ on θ^l, $l = 1, \ldots, 100$ is proportional to

$$p^\theta_{h,l} |\det \Sigma_U(\theta^l)|^{-1/2} \exp\left[-\tfrac{1}{2}\frac{H'_k \Sigma_U(\theta^l)^{-1} H_k}{\omega^2 \kappa^2_{h,k}}\right].$$

11. ω^2

Under the conjugate inverse gamma prior with unit median, $(1/2.198)/\omega^2 \sim \chi^2_1$. From (13)-(15), the conditional posterior distribution for ω^2 then is again inverse-gamma and satisfies

$$\frac{1/2.198 + S^2_c + S^2_g + S^2_h}{\omega^2} \sim \chi^2_{1+(n+10+25)(\tilde{q}+1)}$$

where $S^2_c = \sum_{i=1}^n U'_{c,i}(\kappa^2_{c,i}(1 - \lambda^2_{c,i})\Sigma_U(\theta^u_{c,i}))^{-1}U_{c,i}$, $S^2_g = \sum_{j=1}^{25} U'_{g,j}(\kappa^2_{g,j}(1 - \lambda^2_{g,j})\Sigma_U(\theta^u_{g,j}))^{-1}U_{g,j}$ and $S^2_h = \sum_{k=1}^{10} U'_{h,k}(\kappa^2_{h,k}\Sigma_U(\theta^u_{h,k}))^{-1}U_{h,k}$ with $U_{c,i} = X_i - F - \mu_c \iota_1 - \lambda_{c,i} G_{J(i)}$, $U_{g,j} = G_j - \lambda_{g,j} H_{K(j)}$ and $U_{h,k} = H_k$.

12. $\{\kappa_{c,i}\}_{i=1}^n$

From (5) and (13), the conditional posterior weight of $\kappa_{c,i}$ on κ^l, $l = 1, \ldots, 25$ is proportional to

$$p^\kappa_{c,l} \exp\left[-\tfrac{1}{2}\frac{U'_{c,i}\Sigma_U(\theta^u_{c,i})^{-1}U_{c,i}}{\omega^2(\kappa^l)^2(1-\lambda^2_{c,i})}\right](\kappa^l)^{-(\tilde{q}+1)}$$

with $U_{c,i} = X_i - F - \mu_c \iota_1 - \lambda_{c,i} G_{J(i)}$.

13. $\{\kappa_{g,j}\}_{j=1}^{25}$

From (6) and (14), the conditional posterior weight of $\kappa_{g,j}$ on κ^l, $l = 1, \ldots, 25$ is proportional to

$$p^\kappa_{g,l} \exp\left[-\tfrac{1}{2}\frac{U'_{g,j}\Sigma_U(\theta^u_{g,j})^{-1}U_{g,j}}{\omega^2(\kappa^l)^2(1-\lambda^2_{g,j})}\right](\kappa^l)^{-(\tilde{q}+1)}$$

with $U_{g,j} = G_j - \lambda_{g,j} H_{K(j)}$.

14. $\{\kappa_{h,k}\}_{k=1}^{10}$

From (7) and (15), the conditional posterior weight of $\kappa_{h,k}$ on κ^l, $l = 1, \ldots, 25$ is proportional to

$$p^\kappa_{h,l} \exp\left[-\tfrac{1}{2}\frac{H'_k\Sigma_U(\theta^u_{h,k})^{-1}H_k}{\omega^2(\kappa^l)^2}\right](\kappa^l)^{-(\tilde{q}+1)}.$$

15. $\{J(i)\}_{i=1}^{n}$

    From (5), the conditional posterior weight of $J(i)$ on $j = 1, \dots, 25$ is proportional to

$$\exp[-\tfrac{1}{2}(C_i - \mu_c \iota_1 - \lambda_{c,i} G_j)'(\omega^2 \kappa_{c,i}^2 (1 - \lambda_{c,i}^2) \Sigma_U(\theta_{c,i}^u))^{-1}(C_i - \mu_c \iota_1 - \lambda_{c,i} G_j)].$$

16. $\{K(j)\}_{i=1}^{25}$

    From (6), the conditional posterior weight of $K(j)$ on $k = 1, \dots, 10$ is proportional to

$$\exp[-\tfrac{1}{2}(G_j - \lambda_{g,j} H_k)'(\omega^2 \kappa_{g,j}^2 (1 - \lambda_{g,j}^2) \Sigma_U(\theta_{g,j}^u))^{-1}(G_j - \lambda_{g,j} H_k)].$$

17. (f_0, μ_m)

    Let $\iota_{1:2}$ be the first two columns of $I_{\tilde{q}+1}$, so that $\iota_1 f_0 + \iota_2 \mu_m = \iota_{1:2}(f_0, \mu_m)'$. Then from (8), the conditional posterior of (f_0, μ_m) under the improper flat prior is

$$(f_0, \mu_m)' \sim \mathcal{N}(m_f, V_f)$$

    where $V_f = (\iota_{1:2}' \Sigma_F^{-1} \iota_{1:2})^{-1}$ and $m_f = V_f \iota_{1:2}' \Sigma_F^{-1} F$, with $\Sigma_F = \sigma_m^2 \Sigma_m(\rho_m) + \sigma_{\Delta a}^2 \Sigma_a$.

18. $\sigma_{\Delta a}^2$

    Under the conjugate inverse gamma prior with median 0.03^2, $(0.03^2/2.198)/\sigma_{\Delta a}^2 \sim \chi_1^2$. From (11) the posterior conditional distribution then satisfies

$$\frac{(0.03^2/2.198) + A' \Sigma_a^{-1} A}{\sigma_{\Delta a}^2} \sim \chi_{1+\tilde{q}+1}^2$$

    with $A = F - \iota_1 f_0 + \iota_2 \mu_m - S_m$.

19. σ_m

    From (10), the conditional posterior weight of σ_m on σ^l, $l = 1, \dots, 25$ is proportional to

$$p_l^{\sigma_m} \exp[-\tfrac{1}{2}(\sigma^l)^{-2} S_m' \Sigma_m(\rho_m)^{-1} S_m](\sigma^l)^{-(\tilde{q}+1)}$$

    where $p_l^{\sigma_m}$ are the triangular prior weights.

20. ρ_m

    From (10), the conditional posterior weight of ρ_m on ρ^l, $l = 1, \dots, 25$ is proportional to

$$\exp[-\tfrac{1}{2} S_m' (\sigma_m^2 \Sigma_m(\rho^l))^{-1} S_m] |\det \Sigma_m(\rho^l)|^{-1/2}.$$

21. p_c^λ

From the conjugate nature of the Dirichlet-multinomial prior, the conditional posterior for $\{p_{c,l}^\lambda\}_{l=1}^{25}$ satisfies

$$\{p_{c,l}^\lambda\}_{l=1}^{25} \sim \{W_l / \sum_{i=1}^{25} W_i\}_{l=1}^{25}$$

where W_l are independent χ^2 random variables with $\sum_{i=1}^{n} \mathbf{1}[\lambda_{c,i} = \lambda^l] + 20/25$ degrees of freedom.

22. p_g^λ

From the conjugate nature of the Dirichlet-multinomial prior, the conditional posterior for $\{p_{g,l}^\lambda\}_{l=1}^{25}$ satisfies

$$\{p_{g,l}^\lambda\}_{l=1}^{25} \sim \{W_l / \sum_{i=1}^{25} W_i\}_{l=1}^{25}$$

where W_l are independent χ^2 random variables with $\sum_{j=1}^{25} \mathbf{1}[\lambda_{g,j} = \lambda^l] + 20/25$ degrees of freedom.

23. p_c^θ

From the conjugate nature of the Dirichlet-multinomial prior, the conditional posterior for $\{p_{c,l}^\theta\}_{l=1}^{100}$ satisfies

$$\{p_{c,l}^\theta\}_{l=1}^{100} \sim \{W_l / \sum_{i=1}^{100} W_i\}_{l=1}^{100}$$

where W_l are independent χ^2 random variables with $\sum_{i=1}^{n} \mathbf{1}[\theta_{c,i}^u = \theta^l] + 20/100$ degrees of freedom.

24. p_g^θ

From the conjugate nature of the Dirichlet-multinomial prior, the conditional posterior for $\{p_{g,l}^\theta\}_{l=1}^{100}$ satisfies

$$\{p_{g,l}^\theta\}_{l=1}^{100} \sim \{W_l / \sum_{i=1}^{100} W_i\}_{l=1}^{100}$$

where W_l are independent χ^2 random variables with $\sum_{j=1}^{25} \mathbf{1}[\theta_{g,j}^u = \theta^l] + 20/100$ degrees of freedom.

25. p_h^θ

From the conjugate nature of the Dirichlet-multinomial prior, the conditional posterior for $\{p_{h,l}^\theta\}_{l=1}^{100}$ satisfies

$$\{p_{h,l}^\theta\}_{l=1}^{100} \sim \{W_l / \sum_{i=1}^{100} W_i\}_{l=1}^{100}$$

where W_l are independent χ^2 random variables with $\sum_{k=1}^{10} \mathbf{1}[\theta_{h,k}^u = \theta^l] + 20/100$ degrees of freedom.

26. p_c^κ

From the conjugate nature of the Dirichlet-multinomial prior, the conditional posterior for $\{p_{c,l}^\kappa\}_{l=1}^{25}$ satisfies

$$\{p_{c,l}^\kappa\}_{l=1}^{25} \sim \{W_l / \sum_{i=1}^{25} W_i\}_{l=1}^{25}$$

where W_l are independent χ^2 random variables with $\sum_{i=1}^{n} \mathbf{1}[\kappa_{c,i} = \kappa^l] + 20/25$ degrees of freedom.

27. p_g^κ

From the conjugate nature of the Dirichlet-multinomial prior, the conditional posterior for $\{p_{g,l}^\kappa\}_{l=1}^{25}$ satisfies

$$\{p_{g,l}^\kappa\}_{l=1}^{25} \sim \{W_l / \sum_{i=1}^{25} W_i\}_{l=1}^{25}$$

where W_l are independent χ^2 random variables with $\sum_{j=1}^{25} \mathbf{1}[\kappa_{g,j} = \kappa^l] + 20/25$ degrees of freedom.

28. p_h^κ

From the conjugate nature of the Dirichlet-multinomial prior, the conditional posterior for $\{p_{h,l}^\kappa\}_{l=1}^{25}$ satisfies

$$\{p_{h,l}^\kappa\}_{l=1}^{25} \sim \{W_l / \sum_{i=1}^{25} W_i\}_{l=1}^{25}$$

where W_l are independent χ^2 random variables with $\sum_{k=1}^{10} \mathbf{1}[\kappa_{h,k} = \kappa^l] + 20/25$ degrees of freedom.

Appendix 3 Additional Results

- Figures A1 and A2 referred to in text: **Fig_A1_A2.pdf**
- Country-specific results for the posterior summarized in Table 3 are given in the file **Table3_Results_by_Country.xlsx.**
- A summary of the 50- and 100-year ahead predictive distributions by country are given in the file **Predictive_Distributions_by_Country.xlsx**..
- Posterior Mean CrossCorrelations of $c_{it} - c_{it-50}$:
  - Discussion and network graphs: **CrossCorrelation.pdf**
  - Cross correlations for each pair of countries: **CrossCorrelation_del50_C.xlsx**.
- Figure 4: by country are collected in the file **Figure4_by_Country.zip**
- Figure 6_b: by country are collected in the file **Figure6b_by_Country.zip**.
- Summary of Results for Y/L (output per worker): **YL_Results.pdf**
- POOS forecasting experiment summarized in Section 6.4: **POOS_Experiment.pdf**

Additional References

Bolt, J., R. Inklarr, H. de Jong, and J.L. van Zanden (2018), “Rebasing ’Maddison’: New Income Comparisons and the Shape of Long-Run Economic Development,” Groningen Growth and Development Centre Research Memorandum 174.

Feenstra, R.C., R. Inklarr, and M.P. Timmer (2015), “The Next Generation of the Penn World Tables,” *American Economic Review*, 105(10), pp. 3150-3182.

Geweke, J (2004), “Getting It Right: Joint Distribution Tests of Posterior Simulators,” *Journal of the American Statistic Association*, 99, pp. 799-804.

Inklar, R. and P. Woltjer (2019), “What is new in the PWT 9.1?” manuscript, Groningen Growth and Development Centre.

Phillips, P.C.B (1998), “New Tools for Understanding Spurious Regression,” *Econometrica*, 66, pp. 1299-1325.