NBER WORKING PAPER SERIES

IS GROWTH ADDITIVE?

Callum J. Jones
David López-Salido
Thomas Philippon

Working Paper 35415
http://www.nber.org/papers/w35415

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2026

This paper supersedes and replaces Philippon (2022). We are grateful to Jesús Fernández Villaverde, Tim Cogley, Chad Jones, Ufuk Akcigit, Virgiliu Midrigan, Greg Mankiw, Olivier Blanchard, Xavier Gabaix, Xavier Jaravel, Antonin Bergeaud, Remy Lecat, David Weil, Gilbert Cette, David Romer, German Gutierrez, Ben Jones, Alexey Guzey, Peter Kruse-Andersen, and Bill Easterly for their comments, and to Yad Selvakumar and Nicholas Zevanove for outstanding research assistance. The views expressed are those of the authors and not necessarily those of the Federal Reserve Board, the Banco de España, or the National Bureau of Economic Research.

Is Growth Additive?
Callum J. Jones, David López-Salido, and Thomas Philippon
NBER Working Paper No. 35415
July 2026
JEL No. C10, E17, N0, O11

ABSTRACT

Growth theory is based on the assumption of exponential total factor productivity (TFP) growth, or that the conditional expectation of the next TFP increment is proportional to the current level of TFP. In the U.S., we find strong evidence that TFP growth is conditionally additive, not exponential. Even starting from low priors, Bayesian estimation selects the additive model over the exponential one. In addition, professional forecasts and international TFP series are consistent with the additive model but not with the exponential one.

Callum J. Jones
Federal Reserve Board of Governors
jonescallum@gmail.com

David López-Salido
Banco de España
and CEPR
david.lopez-salido@bde.es

Thomas Philippon
New York University
Leonard N. Stern School of Business
and NBER
tphilipp@stern.nyu.edu

1 Introduction

We propose an empirical investigation of the stochastic process that governs total factor productivity (TFP). At least since Solow (1956) economists have assumed that TFP follows an exponential process (henceforth, model $\mathcal{G}$ for "geometric") which takes the form:

$$\log A_{t+\tau} = \log A_t + g_t\tau + noise, \tag{1}$$

where A_t is TFP in year t and the trend component g_t is constant or at least highly persistent. We will show instead that growth is additive and that the TFP process is better described by model $\mathcal{A}$ (as in "additive" or "arithmetic"):

$$A_{t+\tau} = A_t + b_t\tau + noise, \tag{2}$$

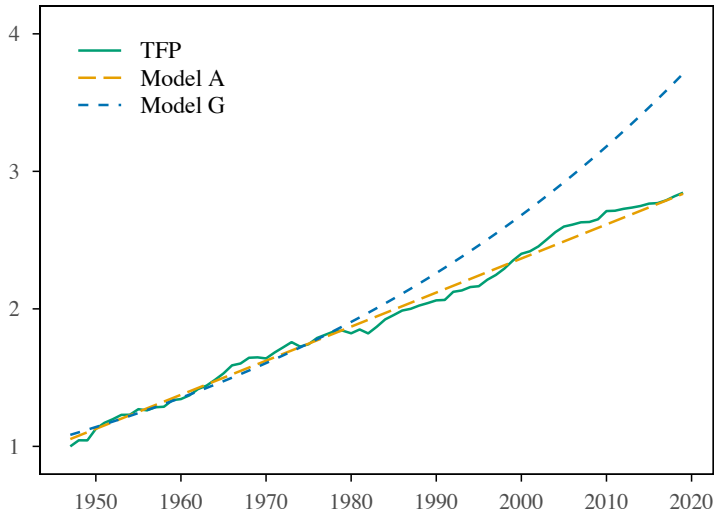
where b_t is constant or at least highly persistent. The key prediction of model $\mathcal{G}$ is that the expected size of the *next* TFP increment is proportional to the *current level* of TFP. We examine data across many countries and time periods and find that this prediction is rejected. In most cases productivity growth appears to be additive.

The exponential specification became standard in growth theory for its theoretical convenience. For example, it delivers a balanced-growth path with stable normalized variables. This analytically convenient benchmark, however, is not an empirical finding that measured TFP increments scales one-for-one with the current level of TFP. We ask whether that benchmark is a good description of measured TFP dynamics.

Figure 1 provides an informal but transparent motivation for our paper. It suggests that TFP growth has been nearly linear in the US since at least World War II. In the figure, models $\mathcal{A}$ and $\mathcal{G}$ are estimated over the first half of the sample (1947-1983) and then used to predict the level of TFP in the second half of the sample (1984-2019). We observe the well-known TFP slowdown "puzzle" with model $\mathcal{G}$, which simply says that actual TFP has fallen short of the exponential benchmark. By contrast there is no TFP slowdown according to model $\mathcal{A}$. To see this, equivalently, Figure 2 shows that the additive growth model predicts the correct path of TFP slowdown.

Empirically, then, US growth after World War 2 is well described by the following statement: Hicks-neutral TFP, normalized to 1 in 1947, increases each year by about 250 basis points. The *initial* trend growth rate is 2.5% but growth is additive: as TFP doubles after 40 year and increments are constant, the measured trend growth

Figure 1: US Post War TFP



Notes: TFP is normalized to 1 in 1947. Models are estimated over 1947-1983. The forecast 1984-2019 is out-of-sample. Data source: Bergeaud et al. (2016). described in Appendix A.

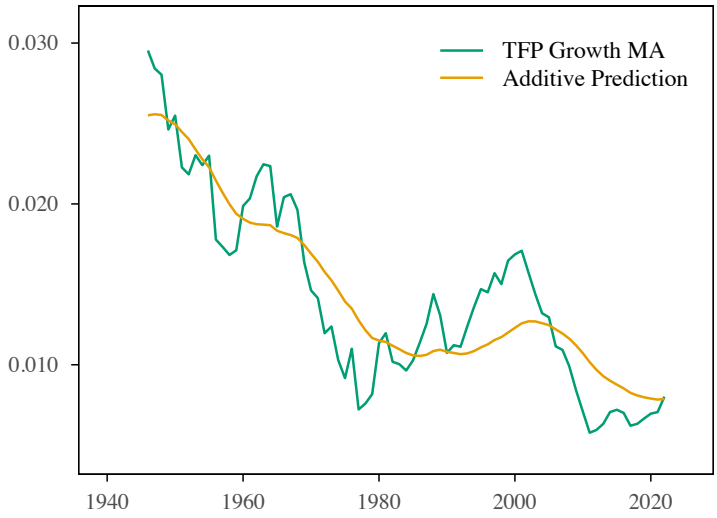
rate in percent is half of what it used to be. After 60 years, it is around one percent, in line with the data.

The point that TFP is additive does not depend on a (possibly noisy) measure of the capital stock. Even when Hicks-neutral TFP grows linearly, capital accumulation creates a convex path for labor productivity. The Appendix shows that the linear TFP model predicts the correct non-linear evolution of labor productivity while the exponential model over-predicts future levels of labor productivity, just like it does for TFP.

The goal of our paper, then, is to formally test the proposition that TFP growth is additive, across many countries and time periods. A challenge, as clearly seen in equations (1) and (2), is to specify the "noise" and obtain estimates of the stochastic trends g_t and b_t. Section 2 presents our statistical framework.

Findings Section 3 focuses on US data and estimates the posterior distributions of models $\mathcal{A}$ and $\mathcal{G}$. We then filter the unobserved states g_t and b_t with a Kalman filter, before approximating the conditional expectations using Monte Carlo simulations. The estimates reveal that the trend in Figure 1 actually starts around 1930. This finding is consistent with the historical literature (Field, 2003; David, 1990; Gordon, 2016), and it suggests that US TFP growth has been additive with approximately

Figure 2: Predicted TFP Slowdown



Notes: TFP Growth MA is the centered moving average of TFP growth over $(t-5, t+5)$. The prediction of model $\mathcal{A}$ is based on a constant annual increment of 250 basis points. Data source: Bergeaud et al. (2016).

constant expected increments for 90 years. Using these estimates, Section 4 conducts a Bayesian model comparison exercise. The exercise confirms the visual impression of Figure 1.

Model $\mathcal{G}$ performs poorly even when the trend g is allowed to change over time, either continuously or with discrete breaks. Thus, in the model selection exercise, the posterior model weight converges to one for model $\mathcal{A}$ by 2022, the end of our sample, even when one starts with large priors in favor of model $\mathcal{G}$ in 1890. In section 6, we report that these conclusions hold across different estimation approaches as well as alternative specifications of models $\mathcal{A}$ and $\mathcal{G}$. In particular, we show that the posterior weight of model $\mathcal{A}$ still converges to one when the trend growth process is autoregressive as opposed to a random walk, and also when we allow the trend growth process to instead be subject to regime breaks at estimated periods. We also show that the results hold under the MLE benchmark and an estimation with a recursively expanding sample.

Section 5 collects two additional exercises that support, but are not central to, the main U.S. evidence. First, we compare the models with long-horizon forecasts from the Council of Economic Advisors (CEA) and the Survey of Professional Forecasters (SPF). Second, we repeat the model-comparison exercise in international TFP data.

The international evidence is useful for external validation, but we interpret it cautiously because many countries in these samples are shaped by catch-up dynamics rather than frontier TFP growth.

A Word of Caution Using a proper Bayesian framework is essential to avoid mistakes and confusions. A first source of confusion concerns the role of conditioning. It is self-evident that all models of growth need time-invariant parameters: productivity growth was barely measurable before 1600, and modest between 1600 and 1800 before accelerating during the industrial revolution, and then again around 1930 (Bouscasse et al., 2025, DeLong, 2022). The drivers of growth were also different: from improved crop yields to a reduction in the importance of land in production, from early machines to industrial research labs, from artisans to global corporations. Models with constant g_t or b_t are non-starters. The only relevant question is this: conditional on the state of the economy at time t, what is the best forecast for the evolution of TFP over the following 10, 20 or 30 years? There is of course no reason to think that the recent history of TFP should provide sufficient information for such a forecast. Our finding that model $\mathcal{A}$ not only beats model $\mathcal{G}$ but also outperforms professional forecasts is then all the more surprising.

A second source of confusion concerns the role of ex-post information. After all, we *know* that there is a slowdown according to the (mis-specified) model $\mathcal{G}$. One may then wonder: would a model with additive (or sub-exponential) growth mechanically outperform model $\mathcal{G}$? The answer is a clear no, but it may useful to explain why even though the argument is rather straightforward. Part of the fallacy comes from ignoring the role of conditioning: as we have just discussed, models with constant g_t or b_t are not even on the table. If the data generating process is conditionally geometric, our Bayesian tests would correctly conclude that $\mathcal{G}$ is the correct model even in the presence of a slowdown. What is true, of course, is that structural breaks, just like any noise, make it more difficult to distinguish among models. So it is possible for the analysis to become inconclusive. Perhaps more surprisingly, we show in Section 4.3 that, in our sample, a further slowdown actually *improves* the posterior of model $\mathcal{G}$. This is because the fit of model $\mathcal{A}$ is quite good, while that of model $\mathcal{G}$ is quite poor. A slowdown (in this case, removing labor quality improvement from measured TFP) hurts the performance of both models, but the impact is relatively larger on model $\mathcal{A}$.

Literature This paper supersedes Philippon (2022), which introduced the additive growth hypothesis and discussed its historical and theoretical motivations. The present paper presents a framework to analyze the empirical plausibility of various stochastic processes that have been proposed to describe the evolution of TFP over time: it develops the state-space estimation framework, estimates posterior distributions for several additive and geometric specifications, conducts Bayesian model comparison for U.S. TFP data, and discusses supporting evidence and robustness exercises.

Our paper uses a Bayesian framework to derive time-varying weights on candidate growth models that adjust as new data arrives. The approach relates to that in Del Negro et al. (2016), which tackles model misspecification by recognizing that no single model perfectly captures economic dynamics; instead, it distributes credibility across models and shifts weights when evidence suggests changes in predictive performance. By updating weights through model forecasting performance, the approach uses Bayesian inference to manage uncertainty and improve robustness.

Our paper also relates with Edge et al. (2007) that examine how private-sector forecasts influence perceptions of long-run U.S. productivity growth and its macroeconomic implications. In the context of a standard geometric growth model, they highlight that shifts in trend productivity—such as those in the 1970s and 1990s—are hard to detect in real time, often confounded by cyclical noise. As supporting evidence, we compare external long-run forecasts with the predictions of the additive and geometric models. This exercise asks whether professional forecasts contain the sustained convexity implied by the geometric model.

Solow (1956) studies the theoretical properties of the neoclassical growth model and Solow (1957) constructs TFP series from 1909 to 1949. Since then, essentially all models of growth have taken the exponential model as a benchmark. One should note, however, that Solow himself did not take a strong stance on the nature of TFP growth. He merely remarked that he could not detect an obvious change in average productivity growth in his short sample. Our Bayesian estimation rejects the exponential model using the full sample, but, interestingly, the power of the test is lower if we only use Solow's original data because of the structural break in 1930.

This paper is not the first to suggest a departure from exponential growth. Jones (1995), for instance, includes a TFP equation of the type $\dot{A}_t = A_t^{\phi} L_{A,t}$ where $L_{A,t}$ is research employment. Exponential growth in standard models comes from the

assumption that $\phi = 1$. The endogenous growth accounting literature calibrates $\phi < 1$ to match the fact that increasing research effort does not necessarily lead to faster growth, but the estimates of ϕ using R&D data are rather unstable. Our results suggest that ϕ is close to zero in the aggregate (not necessarily at the firm level as we discuss in conclusion). Jones (2009) argues that innovation is getting harder because new generations of innovators face an increasing educational burden, while Akcigit and Ates (2023) argue that increasing barriers to knowledge diffusion across firms can explain the evolution of business entry, exit and income shares. Bloom et al. (2020) present case studies of several technologies to argue that innovations are becoming harder to find. Guzey et al. (2021), however, show that this conclusion is sensitive to the choice of a productivity measure, and that many series, including US TFP, do not appear to exhibit exponential growth.

Finally this paper relates to the history of long run growth. The fact that growth increments increase during industrial revolutions speaks to the complementarity of new inventions with existing technologies emphasized by Comin et al. (2010). The turning point of the 1930s is consistent with Field (2003)'s argument that "*the years 1929–1941 were, in the aggregate, the most technologically progressive of any comparable period in U.S. economic history.*" Fernandez-Villaverde et al. (2026) analyze the growth of China in a neoclassical model with time-varying technology transfer: $A_t = \lambda_t A_t^*$, where A is U.S. TFP, A^* is Chinese TFP and λ_t is China's technology level relative to the U.S. at time t. They show that the rapid catch-up of Chinese TFP relative to U.S. TFP is consistent with previous East-Asian miracles, in particular that of Taiwan and Korea. The two approaches are complementary. They analyze the dynamics of capital accumulation, taking as given U.S. TFP. Since they model directly λ_t, the nature of U.S. TFP growth does not matter much for their analysis. By contrast, we do not analyze capital accumulation, but test directly the nature of TFP growth.

2 Modeling TFP Growth

The challenge is to create a test that can distinguish between additive and geometric growth. Let us start with an informal example to build some intuition. Suppose that measured TFP is equal to fundamental TFP A_t^* plus noise $A_t = A_t^* + \epsilon_t$ where ϵ_t is *iid* with volatility σ_ϵ. Let us normalize $A_0^* = 1$, assume that fundamental TFP is

deterministic and that we know the trend g at time 0. We do not know whether A_t^* will grow exponentially as $A_t^* = e^{gt}$, or linearly as $A_t^* = 1 + gt$. How long would we need to wait before deciding which is the correct model? If we want the two forecasts to be $n\sigma$ apart we need $e^{gt} - 1 - gt \geq n\sigma$ or $gt \geq x$ where x is the root of $e^x - 1 - x = n\sigma$. For small values of $n\sigma$, x is close to $\sqrt{2n\sigma}$ so we need $t \geq \frac{\sqrt{2n\sigma}}{g}$. With $n = 2$, $\sigma = 1\%$ and $g = 2\%$ we get approximately $t \geq 10$ years. If we assume instead that the noise is multiplicative $A_t = A_t^* (1 + \epsilon_t)$ then we need approximately $t \geq 12$ years. This back-of-the-envelope calculation gives a sense of the "useful" forecast horizon to test convexity of the TFP growth process.

Let us now proceed with a formal Bayesian analysis.

2.1 Benchmark Specification

We consider two benchmark models for the evolution of observed TFP A_t. We specify model $\mathcal{A}$ (additive growth) as

$$\begin{aligned} A_t &= A_{t-1} + b_t + \sigma_a \epsilon_t^a \\ b_t &= b_{t-1} + \sigma_u u_t \end{aligned}$$

where u_t and ϵ_t^a are independent standard normal shocks. We specify model $\mathcal{G}$ (geometric growth) as

$$\begin{aligned} A_t &= A_{t-1} (1 + g_t) \exp(\sigma_g \epsilon_t^g) \\ g_t &= g_{t-1} + \sigma_\nu \nu_t \end{aligned}$$

where ν_t and ϵ_t^g are independent standard normal shocks. In particular $\log A_t$ is additive, as in equation (1). In the Robustness section below we consider various extensions for the U.S. data such as treating b_t and g_t instead as parameters and allowing for regime shifts in those parameters.

2.2 State space representation

We write each model $m = \mathcal{A}, \mathcal{G}$ in standard VAR(1) representation as

$$x_t^m = \mathbf{Q}_m x_{t-1}^m + \mathbf{G}_m \epsilon_t^m. \tag{3}$$

For $m = \mathcal{A}$ we have $A_t = A_{t-1} + b_{t-1} + \sigma_u u_t + \sigma_a \epsilon_t^a$, so that

$$x_t^{\mathcal{A}} \equiv \begin{bmatrix} A_t \\ b_t \end{bmatrix}, \quad \epsilon_t^{\mathcal{A}} \equiv \begin{bmatrix} \epsilon_t^a \\ u_t \end{bmatrix}, \quad \mathbf{Q}_{\mathcal{A}} \equiv \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{G}_{\mathcal{A}} \equiv \begin{bmatrix} \sigma_a & \sigma_u \\ 0 & \sigma_u \end{bmatrix}.$$

For $m = \mathcal{G}$ we have $\log A_t = \log A_{t-1} + g_{t-1} + \sigma_\nu \nu_t + \sigma_g \epsilon_t^g$ and $g_t = g_{t-1} + \sigma_\nu \nu_t$, so that

$$x_t^{\mathcal{G}} = \begin{bmatrix} \log A_t \\ g_t \end{bmatrix}, \quad \epsilon_t^{\mathcal{G}} = \begin{bmatrix} \epsilon_t^g \\ \nu_t \end{bmatrix}, \quad \mathbf{Q}_{\mathcal{G}} \equiv \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{G}_{\mathcal{G}} \equiv \begin{bmatrix} \sigma_g & \sigma_\nu \\ 0 & \sigma_\nu \end{bmatrix}.$$

With the two models written in their state space forms, we can then use likelihood-based state-space methods to estimate the parameters of the two models.[1] We do this in the next section using Bayesian estimation.

3 Estimation for the U.S.

3.1 Data

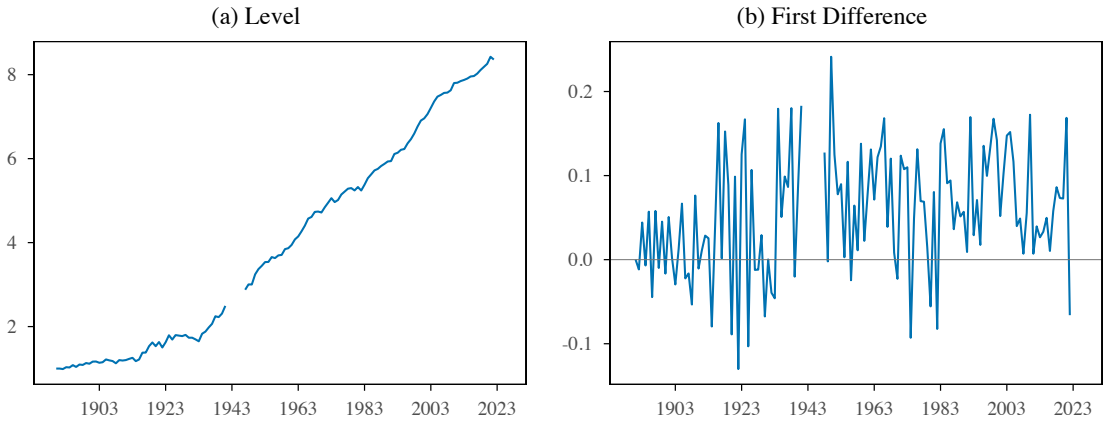
Data on TFP for the U.S. from Bergeaud et al. (2016) is available from 1890 to 2022. We omit the WW2 years 1942–1946. Appendix A describes the U.S. data in more detail. We add a lower barrier of zero for the stochastic trends b_t and g_t based on theoretical priors (i.e., no fundamental technological regress), imposing the same barrier in estimation, filtering, and smoothing, and we normalize the value of TFP in the first year to one. Figure 3 plots the data we use in both levels and first differences.

3.2 Bayesian Estimation of Second Moments

We estimate the parameters $\theta^{\mathcal{A}} = (\sigma_a, \sigma_u)$ for the additive model and $\theta^{\mathcal{G}} = (\sigma_g, \sigma_\nu)$ for the geometric model by combining the Kalman filter likelihood with uniform priors over these parameters and sampling from the resulting posterior distributions. The posterior draws are used in the Bayesian model analyses below. We describe the MCMC algorithm used to sample the posterior in Appendix B.4.

[1] We report the maximum-likelihood benchmark, recursive MLE estimates, autoregressive trend specifications, and the specification where we treat b_t and g_t as estimated parameters subject to regime shifts in periods that are also estimated in Section 6 and Appendix B.

Figure 3: Estimation Sample, U.S. TFP



Notes: The left panel shows the level of TFP from BCL, which is normalized to 1 in 1890. The right panel shows the first difference in TFP. The WW2 years 1942-1946 are omitted from the sample.

Table 1: Bayesian Parameter Estimates (US, 1880-2022)

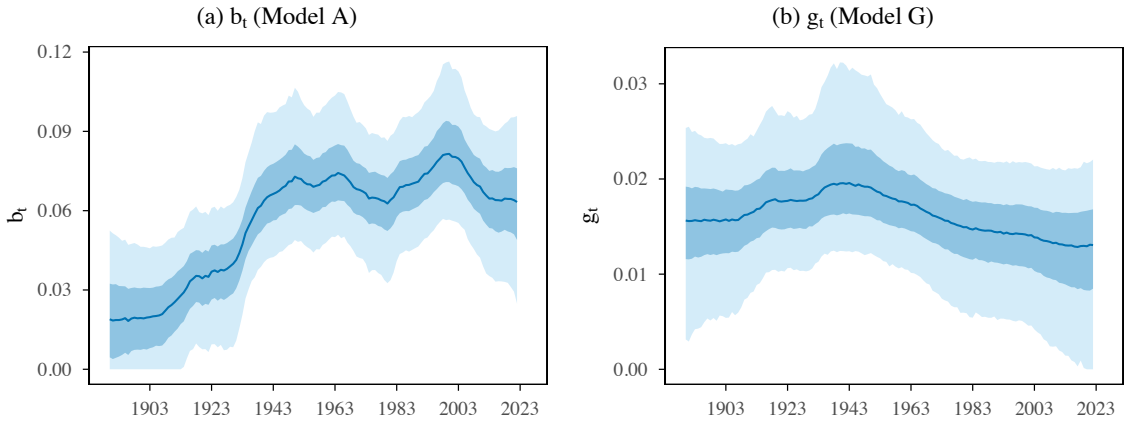
	Model $\mathcal{A}$		Model $\mathcal{G}$	
	σ_u	σ_a	σ_ν	σ_g
Median	0.0061	0.0689	0.0012	0.0325
Mode	0.0047	0.0688	0.0006	0.0321
Std Dev	(0.0036)	(0.0048)	(0.0012)	(0.0022)
5th percentile	0.0025	0.0617	0.0001	0.0292
95th percentile	0.0140	0.0773	0.0037	0.0363

Notes: This table provides posterior summaries for the parameters governing the standard deviations of the $\mathcal{A}$ and $\mathcal{G}$ models. The estimation uses uniform priors over the reported parameters. The 5th and 95th percentiles define the equal-tailed 90 percent credible interval.

Posterior estimates. Table 1 gives medians, modes, standard deviations, and 5th and 95th percentiles for the estimated posterior distributions. Figure B4 in the Appendix shows the posterior densities of the parameters. These posterior summaries are similar to the MLE benchmark estimates reported in Section 6.1.

Posterior inference. For posterior inference, we use 1,000 parameter draws from the posterior distribution. For each parameter draw, we run the constrained Kalman smoother and draw the latent state from the smoothed state distribution subject to the same zero lower barrier. Figure 4 plots the resulting posterior distribution for the smoothed estimates. The median b_t and g_t across those draws is similar to the filtered

Figure 4: Smoothed Estimates of Growth Trends



Notes: This figure shows the smoothed estimates of b_t and g_t for 1,000 draws of the parameters from the posterior distributions, incorporating the conditional state uncertainty from the constrained Kalman smoother and imposing the same zero lower barrier on the trend states. The dark blue areas show the 25th and 75th percentiles of the resulting distribution of b_t and g_t and the light blue areas show the corresponding 5th and 95th percentiles.

paths under the MLE benchmark estimates. We use the same posterior parameter draws for the model comparison exercise in Section 4.

4 Bayesian Model Comparison

This section discusses the statistical approach we take in comparing the two models.

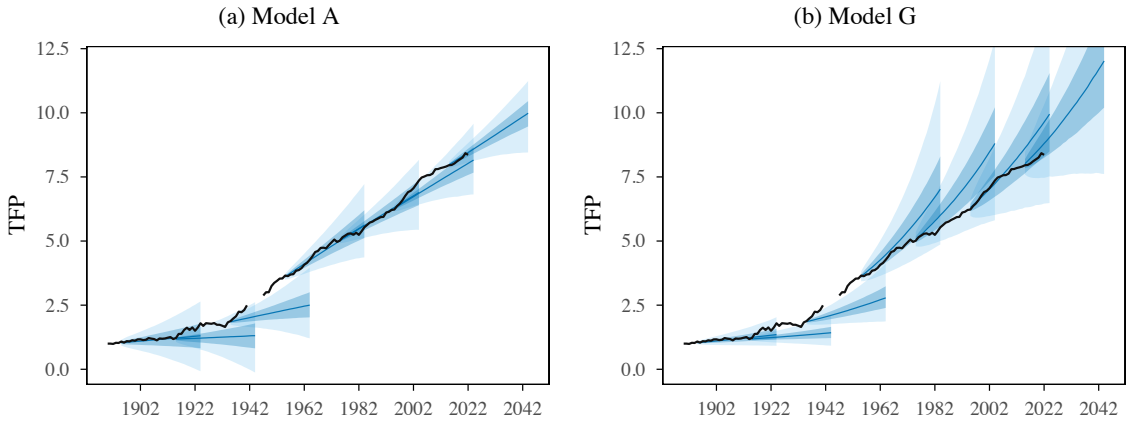
4.1 Cumulative Log Predictive Score

We compare the models by their predictive densities. For the forecast figures in this section, we use the posterior draws from the Bayesian estimation in Section 3.2; Section 6.1 reports the corresponding MLE benchmark. At each forecast origin we compute the predicted density of models $m = \mathcal{A}, \mathcal{G}$:

$$p\left(x \mid A^t, m, h\right) \equiv \Pr\left(A_{t+h} = x \mid A^t, m\right). \tag{4}$$

For both models $\mathcal{A}$ and $\mathcal{G}$, we construct these predictive densities from the posterior distribution. Let θ^s denote a retained posterior draw of the model parameters. For each forecast origin t and each draw θ^s, we first run the Kalman filter using θ^s and

Figure 5: TFP Level and Forecasts



Notes: The blue line plots the model predicted median paths of TFP (every 20 years, over a forecast horizon of 30 years), along with the 5% / 95% range (in light blue) and the 25% / 75% range (in darker blue).

initialize the forecast at the corresponding filtered state. We then compute the h-step-ahead predictive density implied by the model's state-space law of motion under θ^s. Averaging these densities across posterior parameter draws gives the posterior predictive distribution. Thus the forecast bands integrate over both parameter uncertainty and future shock uncertainty, rather than conditioning on posterior modal parameters. Figure 5 displays the results.

Figure 5 is the key figure of the paper. The black line is the actual path of US TFP. The blue lines are the median model forecasts and the shaded areas are the posterior predictive intervals. We will discuss the comparison with the SPF and CEA forecasts in the next section. Of course, neither model can forecast the structural change in trend growth in the 1930s, so both models provide pessimistic forecasts of TFP in the 1950s. As the models filter the early post-war data they revise upward their estimates of b_t and g_t. But the medium / long term forecasts of model $\mathcal{G}$ are widely optimistic, in addition to being very imprecise. Model $\mathcal{A}$, on the other hand, is broadly consistent with the historical evidence. We now formalize and quantify these points.

We evaluate the relative performance of models $\mathcal{A}$ and $\mathcal{G}$ by comparing their out-of-sample forecast performance. Given *realized* productivity A_{t+h} the log predictive score is

$$\log p\left(A_{t+h} \mid A^t, m, h\right),$$

where p is defined in (4). The key statistic used is the *cumulative log score* (CLS), computed as the sum across time of the log predictive scores for h-step-ahead forecasts:

$$\mathrm{CLS}_m(T) = \sum_{t=1}^{T-h} \log p\left(A_{t+h} \mid A^t, m, h\right). \tag{5}$$

We use $h = 10$ years as our benchmark forecast horizon and therefore T runs from 1900 to 2022. Our results are robust to using longer horizons, as shown in Appendix B for $h = 15$ years. We use these scores as a posterior-predictive, or prequential, comparison of forecast performance; we do not assume that either specification is the true data-generating process.

4.2 Log Bayes Factor and Posterior Model Weights

From the cumulative log scores in (5), the log Bayes factor in favor of Model $\mathcal{A}$ is

$$\log B_{\mathcal{A},\mathcal{G}}(T) = \mathrm{CLS}_{\mathcal{A}}(T) - \mathrm{CLS}_{\mathcal{G}}(T). \tag{6}$$

When $\log B_{\mathcal{A},\mathcal{G}}(T) > 0$, Model $\mathcal{A}$ has superior cumulative forecast accuracy; when negative, Model $\mathcal{G}$ does.

Figure 6, Panel (a) shows the log Bayes factors in favor of model $\mathcal{A}$. Panel (a) shows that the data up to 1960 is inconclusive for two main reasons. Firstly, TFP growth before WWI is rather slow so the convexity effect is relatively small. Second, as is well known from historical research (Field, 2003; David, 1990; Gordon, 2016) there is a structural shift in the 1930s with the large scale implementation of several new technologies. The positive change in the slope of the TFP series makes it difficult to differentiate the two models. By the late 1960s, however, the lack of convexity in the TFP series starts to become more obvious to our Bayesian estimator.

Panel (b) of Figure 6 translates the log Bayes factors into posterior probability weights, using an adjustment proposed by Del Negro et al. (2016). Let us start with standard Bayesian Model Averaging (BMA). If we assign prior probabilities $\pi_{\mathcal{A}}$ and $\pi_{\mathcal{G}} = 1 - \pi_A$ to the two models, then the posterior probability of Model $\mathcal{A}$ given sample T using the standard BMA formula is

$$\mathbb{P}_T(\mathcal{A}) = \frac{\pi_A \cdot B_{\mathcal{A},\mathcal{G}}(T)}{\pi_A \cdot B_{\mathcal{A},\mathcal{G}}(T) + \pi_G} = \frac{\pi_A \cdot \exp\left[\mathrm{CLS}_{\mathcal{A}}(T) - \mathrm{CLS}_{\mathcal{G}}(T)\right]}{\pi_A \cdot \exp\left[\mathrm{CLS}_{\mathcal{A}}(T) - \mathrm{CLS}_{\mathcal{G}}(T)\right] + \pi_G}. \tag{7}$$

As Del Negro et al. (2016) point out, however, standard BMA can produce unreasonable swings in posterior probabilities. There are two broad classes of explanations for this problem. First, BMA assumes that exactly one of the two models is correct, which is not plausible in most economic studies, ours included. At most we wish to claim that the linear process provides a useful approximation of the true TFP process and that we can reject the strong convexity of the geometric process, but obviously there are models with weaker convexity that we could not reject. Second, the estimation, including Kalman filtering, assumes normal disturbances. The filtering itself is robust to non-normality since it is an orthogonal projection in a Hilbert space, but the implied posterior probabilities are sensitive to the specification of the error terms and normality gives a false sense of precision: a CLS of ± 3 can easily happen with noisy data, but since $\exp(3) \approx 20$, one would conclude strongly in favor of one model when no such confidence is in fact warranted.

Del Negro et al. (2016) propose instead a dynamic prediction pool and a scaling factor γ that tempers the elasticity of the posterior probability weight to the log Bayes factor. We use the following specification for the posterior probability weight in the dynamic prediction pool:

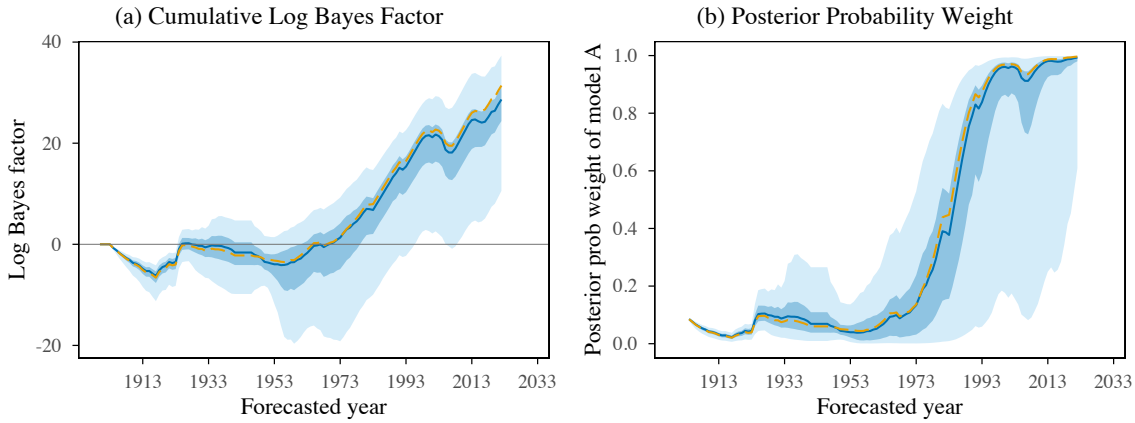
$$w_T = \frac{\pi_A \cdot \exp\left[\gamma \left\{\text{CLS}_{\mathcal{A}}(T) - \text{CLS}_{\mathcal{G}}(T)\right\}\right]}{\pi_A \cdot \exp\left[\gamma \left\{\text{CLS}_{\mathcal{A}}(T) - \text{CLS}_{\mathcal{G}}(T)\right\}\right] + \pi_G}. \tag{8}$$

A small value of γ makes the posterior weight react less sharply to recent differences in forecast accuracy. In our exercises below, we set γ to a conservative value of 0.25. Under this value, simulation exercises show that it takes about 20 years for cumulative log score differences of around 1 each period to raise a posterior model weight from 0.1 to near 1. In the estimation below the main effect of $\gamma = 0.25$ is to lower the volatility of the estimated posterior weight in the 1930s. Clearly the conclusion that the weight ≈ 1 when $T > 1980$ would only be stronger with $\gamma = 1$.

As shown in panel (b) of Figure 6, the posterior weight of model $\mathcal{A}$ under (8) hovers around the prior of 0.1 until about the 1970s. From then on, model $\mathcal{A}$ shows superior forecast accuracy for future TFP, with the cumulative log Bayes factor shifting strongly towards it. This pattern holds both for the median across posterior draws and for the pooled posterior-predictive comparison. Apart from a short dip in the 2000s, the posterior weight of model $\mathcal{A}$ is near 1.[2]

[2]In Appendix B.1, we complement this density-based evaluation with a point-forecast comparison

Figure 6: Log Bayes Factor and Posterior Model Weight



Notes: This figure shows the cumulative log Bayes factor and the posterior probability weight of model $\mathcal{A}$ across posterior parameter draws. For each draw, we filter the state path, compute the 10-year-ahead predictive density implied by the model's state-space law of motion, and then compute the cumulative log score path. The solid line is the median across draws; the dark blue areas show the 25th and 75th percentiles and the light blue areas show the 5th and 95th percentiles. The dashed orange line is the pooled posterior-predictive comparison, which averages predictive densities across posterior draws before accumulating log scores. The initial prior on model $\mathcal{G}$ is 0.9. We use $\gamma = 0.25$ in constructing the posterior probability weight.

4.3 Bayesian Model Comparison and Lower TFP Growth

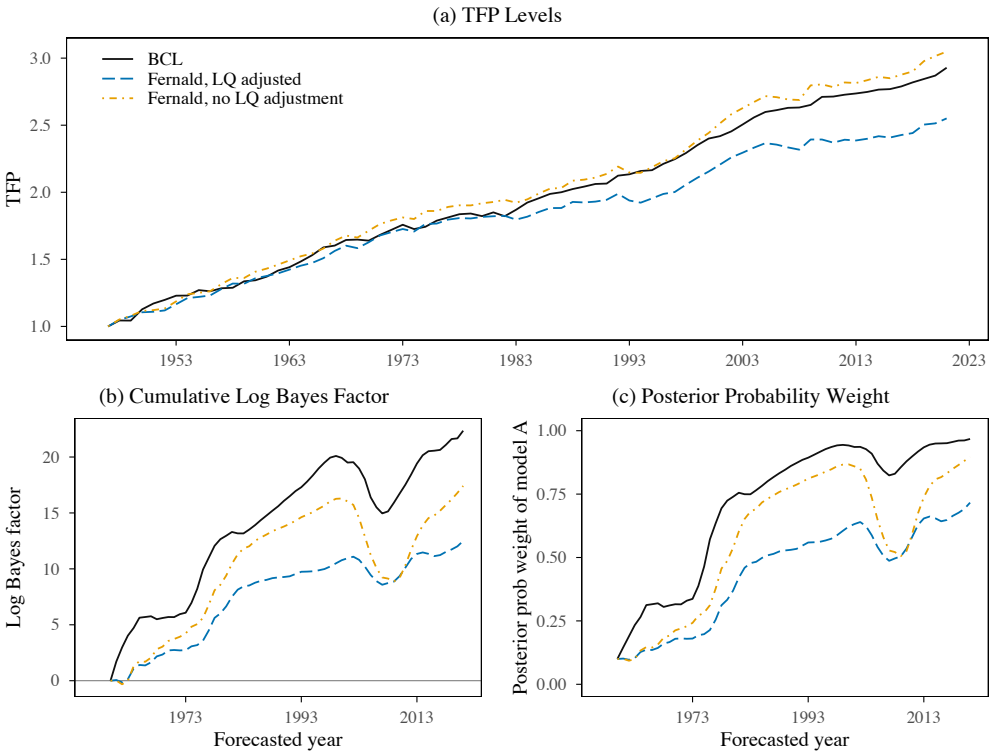
We repeat in Figure 7 the same posterior-predictive comparison using Fernald's post-war utilization adjusted TFP series, both with and without the labor-quality adjustment. This comparison is useful because the labor-quality adjustment removes a quantitatively important source of postwar TFP growth, thus creating a further slowdown.[3]

At first glance, one might expect a lower-growth series to make the model assessment tilt even more strongly toward the additive specification. The results show why that intuition is incomplete. The additive model continues to receive higher posterior-predictive weight in both cases in the Fernald data, but the margin is *smaller* for the slower-growing labor-quality-adjusted series. The reason is that lower growth com-

based on mean squared error. This simpler exercise abstracts from predictive uncertainty. We find that the additive specification continues to outperform the geometric model in MSE terms.

[3]The labor-quality adjustment lowers TFP by about 16% by the end of the sample (2021). We do not advocate in favor of the labor-quality adjustment, we simply use it as a robustness check, and as an illustration for an important point. Whether one should remove labor quality improvements from measured TFP is obviously a *theoretical* question. In many models – in particular models of growth via learning – the answer is no; in some models the answer is yes.

Figure 7: Alternative TFP Measures and Model Fit



Notes: Panel (a) compares TFP levels. Panels (b) and (c) report pooled posterior-predictive comparisons at the 10-year horizon. BCL is re-estimated on the common Fernald sample, with TFP normalized to one in 1947. Fernald no LQ adds the labor-quality contribution back to utilization-adjusted TFP growth.

presses the distance between the additive and geometric forecasts over a 10-year horizon. In particular, removing labor-quality growth produces a flatter postwar TFP path, so the compounded forecasts of the geometric model are penalized less even though the series is generated by an additive process. Equivalently, when trend growth is slower, the 10-year additive and geometric forecasts are closer together, so the cumulative log-score comparison requires a longer sample to separate the two models.

A simulated-data exercise in Appendix A illustrates the point under an additive data-generating process. In the case when there is a unknown break in b, the additive growth increment, to a lower value, the posterior probability weight for model $\mathcal{A}$ is lower by the end of the sample compared to a simulation without the break or a break to a higher value of b.

4.4 Revisiting Solow (1957)

> "Not only is $\Delta A/A$ uncorrelated with K/L, but one might almost conclude from the graph that $\Delta A/A$ is essentially constant in time, exhibiting more or less random fluctuations about a fixed mean."
>
> Solow (1957)

The Bayesian approach to model selection can also shed light on the history of economic research on growth. Using data from 1909 to 1949, Solow (1957) found a pattern for A_t qualitatively similar to that in Figure 11. He wrote that "*there does seem to be a break at about 1930. There is some evidence that the average rate of progress in the years 1909-29 was smaller than that from 1930-49.*" Indeed, a formal test finds a structural break around 1930 in the first difference series $A_t - A_{t-1}$. The change in the slope makes it difficult to distinguish models $\mathcal{A}$ and $\mathcal{G}$ using only Solow's data. Formally, if one feeds data from 1909 to 1949, the posteriors over the two models are not very different from the priors. Thus Solow made a reasonable choice based on available evidence at the time of his writing, and it is only with the benefit of post-1950 data that we can confidently conclude that growth is additive.

5 Additional Evidence

The previous sections provide the main evidence using U.S. TFP data. This section reports two additional exercises. The first asks whether external long-horizon forecasts are closer to the additive or geometric model. The second asks whether the broad pattern is visible outside the U.S.

5.1 Professional Forecasters

Economists have used the geometric model for its theoretical convenience, not for its accuracy. Professional forecasters, on the other hand, have incentives to use whichever model they find most plausible. The logic is similar to Friedman (1953)'s billiard-player argument. The point is not that forecasters compute formal Bayes factors, but that successful forecasting procedures should be, in Friedman's phrase, "capable of reaching essentially the same result": they should avoid systematic convexity if that convexity is a poor description of long-run TFP dynamics.

We use forecasts of labor productivity from the CEA and the SPF to compare the models with external long-horizon forecasts (see the data Appendix A for details of each forecast). We convert them to TFP forecasts assuming a Cobb-Douglas production function. We then evaluate each forecast under the predictive density implied by each posterior parameter draw. For each draw, we use the corresponding filtered state at the forecast origin, compute the log predictive scores of the SPF and CEA forecasts, cumulate the log-score difference, and map it into a posterior probability weight using the same updating rule as in Section 4. This gives a posterior-draw distribution for the model weight implied by the external forecasts.

Figure 8 shows our results. Panel (a) presents the data on CEA and SPF forecasts, alongside the posterior predictive range for TFP from model $\mathcal{A}$, with *forecasted* year on the time axis. We also plot realized TFP in those years in panel (a). SPF forecasts are available from 1992:Q1 to 2022:Q3 at a consistent 10 year horizon. The first star, in 2002, corresponds to the forecast made in 1992. CEA forecasts are available as early as 1970 but with inconsistent time horizons, as we explain in Appendix A.

Panel (a) shows that model $\mathcal{A}$ is broadly consistent with SPF forecasts and with CEA forecasts except for the 2017-2021 period where CEA forecasts appear optimistic in a way which is inconsistent both with model $\mathcal{A}$ and with SPF forecasts. Thus external long-horizon forecasts do not appear to build in much convexity in future TFP. Model $\mathcal{A}$ also provides better forecasts of actual TFP than either SPF or CEA in parts of the sample, especially between 2010 and 2020.

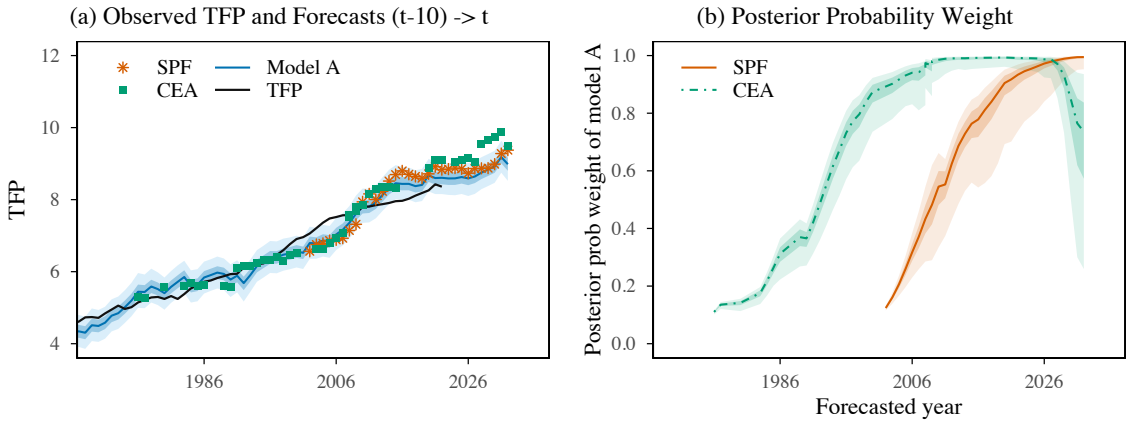
Panel (b) translates the forecasts and CLS into posterior probability weights. The lines are median weights across posterior draws and the shaded areas show the distribution across draws. Applying the same posterior-predictive model-comparison statistic to these forecasts assigns high weight to model $\mathcal{A}$. For CEA forecasts, the weight converges to almost one but decreases with the release of the 2017-2021 forecasts, as CEA economists publish unexpectedly optimistic forecasts of TFP during those years.

5.2 International Evidence

Bergeaud et al. (2016) provide data for 23 countries.[4] We apply the models of Sec-

[4]Australia, Austria, Belgium, Canada, Switzerland, Chile, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Ireland, Italy, Japan, Mexico, Netherlands, Norway, New Zealand, Portugal, Sweden, and the United States. We also use provided data for a Euro Area aggregate. The

Figure 8: Comparison with Forecasters



Notes: Year t is the forecasted year. The shaded areas are the 25% and 75% bands (solid) and 5% and 95% bands (lighter shade) for the posterior predictive distribution of model $\mathcal{A}$, initialized at the filtered state at the forecast origin and integrating over posterior parameter draws and future shocks. Stars and squares are median SPF and CEA forecasts for year t. Note that growth forecasts for the CEA are for different horizons each year (i.e., sometimes less than 10 years), as described in the data Appendix A. In panel (b), lines are medians across posterior draws and shaded areas are 25% and 75% bands (darker shade) and 5% and 95% bands (lighter shade) for the draw-specific posterior probability weights. The initial prior on the geometric model is 0.9 and $\gamma = 0.25$.

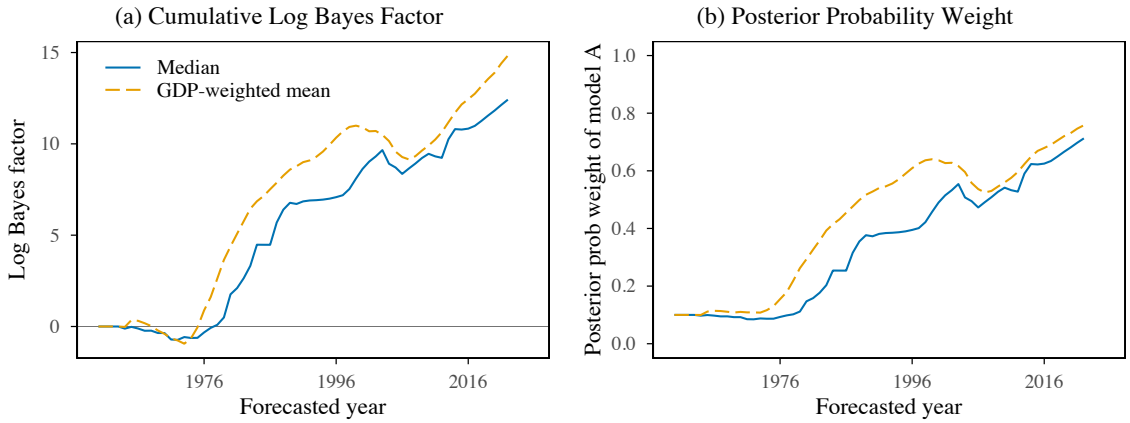
tion 3 to each country as an external validation exercise, estimating country-specific posterior distributions with the same uniform priors and constrained Kalman filters used in the U.S. baseline. These data are informative about whether additive-looking TFP dynamics are widespread, but they are not all equally informative about frontier growth because many countries are affected by catch-up, sectoral reallocation, and measurement differences. With that qualification, model $\mathcal{A}$ beats model $\mathcal{G}$ in almost all countries, and often by a wider margin than in the U.S.

Figure 9 summarizes the international evidence from the BCL dataset by plotting the median and real-GDP-weighted mean across countries of the cumulative log Bayes factors and posterior model weights. The median log Bayes factor becomes positive by the late 1970s, indicating that, on average, the additive model delivers superior 10-year-ahead forecasts across countries. In addition, most countries in the sample show a positive log Bayes factor, with the notable exception of Ireland where GDP-based measures differ significantly from GNI-based measures.[5]

sample covers the post-WW2 period 1950–2022. Appendix C plots the TFP paths for the countries used in the analysis.

[5]We make no attempt to correct the data for well-known issues, such as GNI vs GDP in Ireland, to avoid any risk of bias. We simply use the GDP-based TFP estimated from Bergeaud et al. (2016).

Figure 9: Model Comparison, BCL Dataset



Notes: The lines plot the median across countries in the BCL dataset and the real-GDP-weighted mean across countries. The figure is based on 10-year-ahead posterior-predictive scores computed from country-specific posterior parameter draws. The initial prior on the geometric model is 0.9 and $\gamma = 0.25$.

We next extend the analysis to a selection of fast-growing Asian economies using Penn World Table data: Korea, Taiwan, and China. These cases are especially useful as stress tests because rapid catch-up can generate dynamics that differ from those at the frontier. Figure 10 shows both the TFP paths and the posterior probability weights implied by 10-year-ahead predictive scores. Despite very different initial conditions and episodes of rapid catch-up, TFP paths in Korea and Taiwan are approximately linear in levels over extended periods. In particular, there is little evidence of sustained convexity in TFP, even during phases commonly described as "miracle growth."

We use as the initial prior for the geometric model in each country the posterior probability weight estimated for the U.S. at the date that the forecast is made.[6] For Korea and Taiwan, the posterior probability weights shift toward model $\mathcal{A}$, especially at the 10-year horizon. Taiwan is probably the most striking example of linear TFP growth. The posterior-predictive comparison also shifts toward model $\mathcal{A}$ for China, but we interpret this evidence cautiously. China experiences a structural break around 1980, from stagnation to strongly increasing TFP, and there is some growth acceleration between 2000 and 2010. In that sense Chinese TFP around 1980 is similar to

Individual country results are shown in Appendix C.

[6]Under this, the initial prior probability for the geometric model is around 0.9, which was our assigned prior on the geometric model in the exercises above.

Table 2: MLE Parameter Estimates (US, 1880-2022)

	Model $\mathcal{A}$		Model $\mathcal{G}$	
	σ_u	σ_a	σ_ν	σ_g
MLE Estimate	0.005	0.069	0.001	0.032
Standard Deviation	(0.002)	(0.005)	(0.001)	(0.002)

Notes: This table provides the parameter estimates for the parameters governing the standard deviations of the $\mathcal{A}$ and $\mathcal{G}$ models under MLE.

US TFP around 1930, so the result should be viewed as supportive evidence rather than as a clean frontier-growth test.

6 Extensions and Robustness

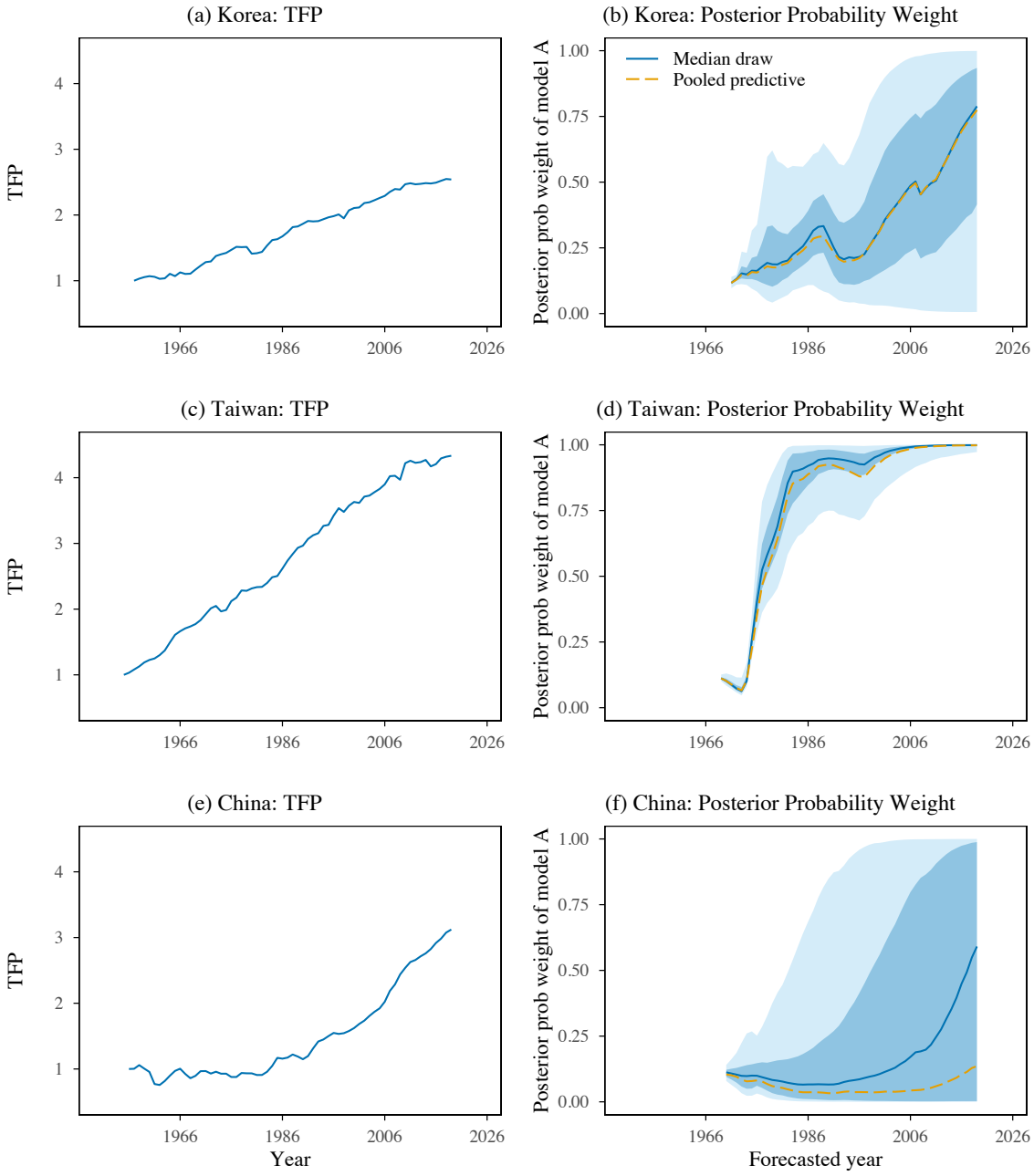
In this section, we report the MLE benchmark estimates used in the predictive-score exercises and consider a number of extensions of the baseline framework using U.S. data. We show that our results are robust to estimating the model in real-time, to alternative priors over the trend innovation variance, and to alternative specifications, including an autoregressive assumption for the trend growth terms as well as treating the trend growth terms as parameters subject to regime breaks.

6.1 Maximum Likelihood Estimates

As a benchmark, we estimate the parameters $\theta^{\mathcal{A}} = (\sigma_a, \sigma_u)$ for the additive model and $\theta^{\mathcal{G}} = (\sigma_g, \sigma_\nu)$ for the geometric model using full-sample maximum likelihood estimation (MLE). Note, however, that we only estimate the second moments of the data in the full sample, so the risk of creating look-ahead bias is small. Nonetheless, as a robustness, we also run a real time MLE estimation over expanding samples and show that the results of our model evaluation exercise are quantitatively similar. Table 2 shows the MLE estimates in our benchmark specification.

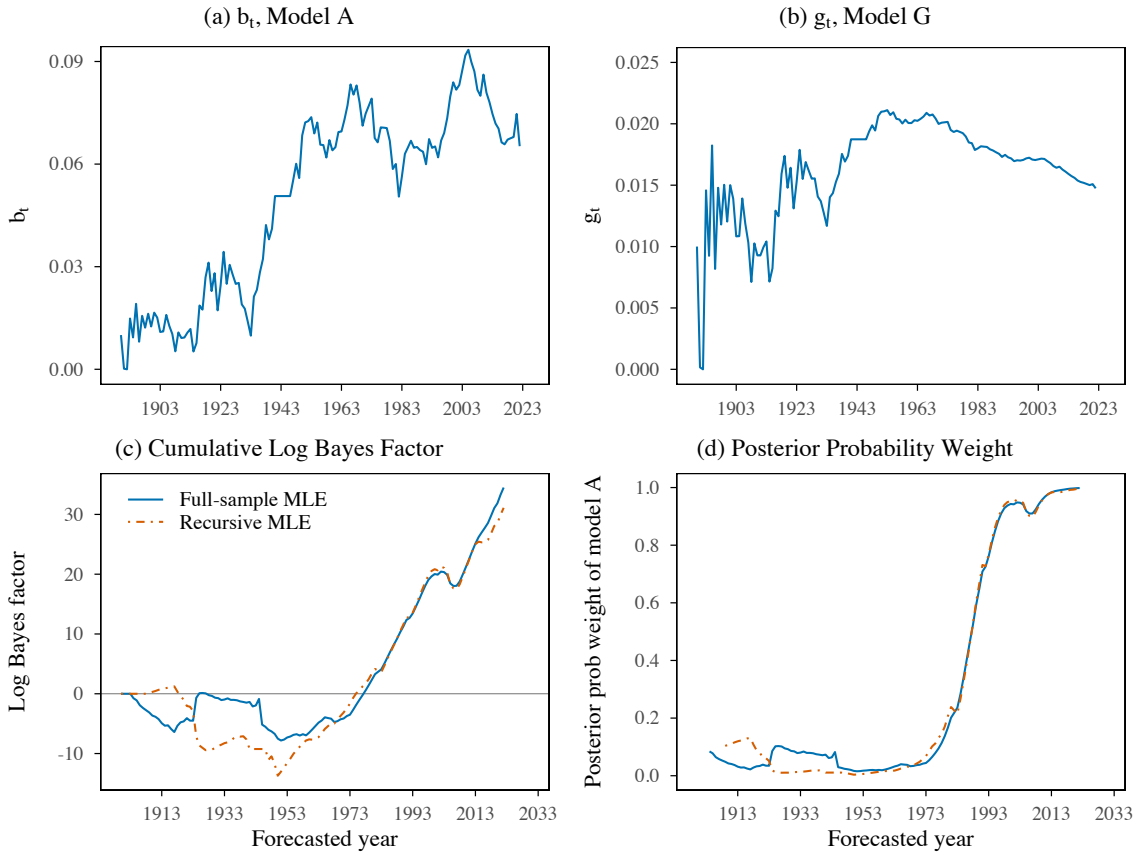
Figure 11 reports the full-sample MLE filtered state estimates and compares the corresponding predictive-score evidence with the recursive MLE exercise. The state estimates fix the parameters at their full-sample MLE values and filter the states using data $A_1, \ldots, A_t$ only. The recursive exercise re-estimates the model parameters at each forecast origin using the expanding sample through t and then recomputes the forecast scores. The value of b_t is around 0.02 during the first part of the sample

Figure 10: TFP and Model Comparison, Select Countries in PWT Dataset



Notes: Panels (a), (c), and (e) plot TFP at constant national prices for Korea and Taiwan, and welfare-relevant TFP at constant national prices for China. Panels (b), (d), and (f) plot posterior probability weights for model $\mathcal{A}$ based on 10-year-ahead posterior-predictive scores computed from country-specific posterior parameter draws. The solid line is the posterior-draw median; shaded areas are 25% and 75% bands (darker shade) and 5% and 95% bands (lighter shade). The dashed line is the pooled posterior-predictive weight. The prior on the geometric model for each country is set to the posterior probability weight calculated for the U.S. in the period that the forecast is made and $\gamma = 0.25$.

Figure 11: MLE Filtered States and Model Evaluation



Notes: Panels (a) and (b) show the filtered b_t and g_t under full-sample MLE parameters. Panels (c) and (d) compare the cumulative log Bayes factor and posterior probability weight of model $\mathcal{A}$ at a 10-year forecast horizon under full-sample MLE parameters and recursive expanding-window MLE estimates. When computing the posterior probability weight, the prior of model $\mathcal{G}$ is 0.9. We use $\gamma = 0.25$ in constructing the posterior probability weight.

up until about the 1930s, after which it is estimated to increase to around 0.07 by 1950 and fluctuate around that value. The value of g_t is estimated to be around 0.015 until about the 1930s, rises to around 0.02 when growth accelerates after the 1930s, and then is estimated to trend down from the 1970s.

For the predictive-score comparison, we initialize forecasts at the filtered state estimates for each forecast origin and create TFP forecasts $\{A_{t+h}\}_{h=1}^{30}$. Finally we use the forecasts to sequentially assess and compare the predictive performance of the models.

6.1.1 Recursive Maximum Likelihood Estimates

The benchmark predictive-score results use the full sample MLE parameter estimates to evaluate the models. In this section, we instead evaluate the models using recursive estimates based on expanding window. That is, each period we expand the dataset by one year, recompute the MLE estimates of the model parameters on this dataset, and construct the log scores and cumulative log Bayes factors using the period-by-period estimates. Panels (c) and (d) of Figure 11 show the results evaluated at a 10-year forecast horizon, with our full sample MLE estimates also plotted for comparison. As the figure shows, by the end of the sample, the $\mathcal{A}$ model is preferred with a posterior model weight ≈ 1, as was the case under the full sample estimates. The recursive estimation also reinforces the point made in section 4.4 about the observations made by Solow (1957) about $\Delta A/A$ being roughly constant over time, with the strongest evidence for geometric growth occurring around 1950.

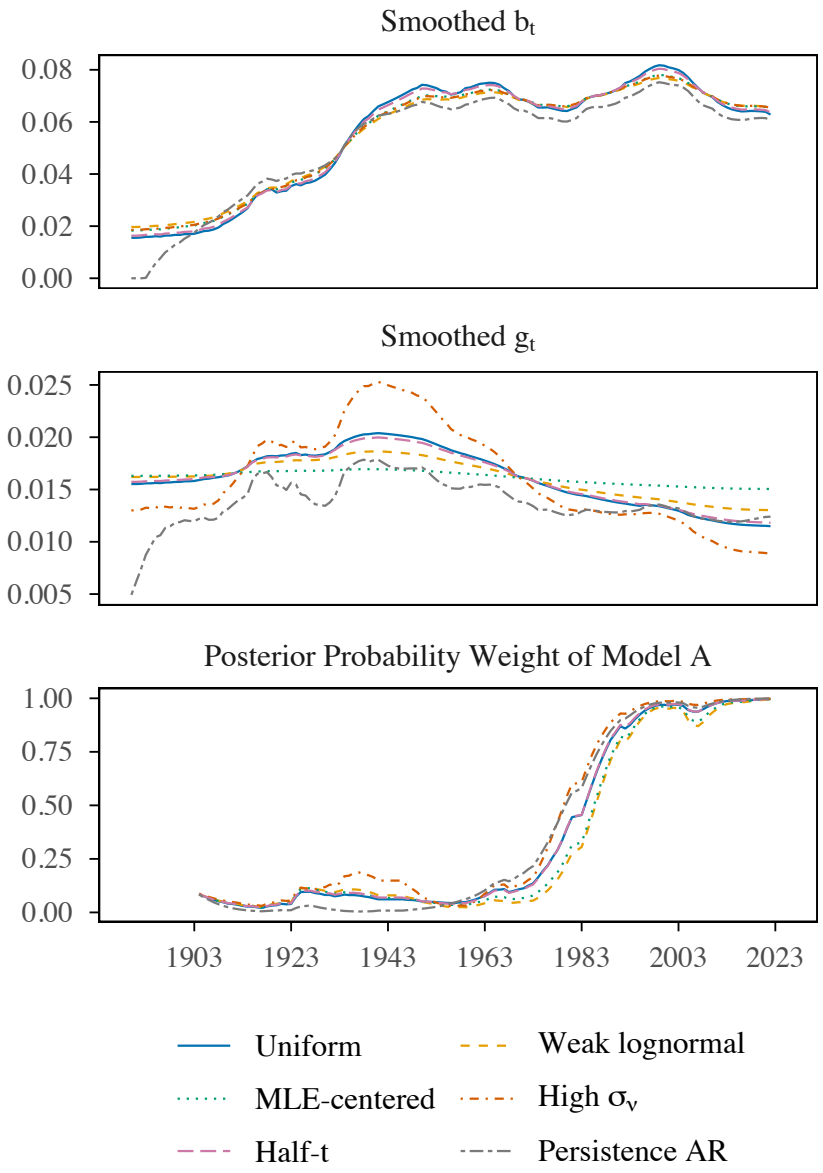
6.2 Prior Robustness

The Bayesian baseline uses deliberately weak priors on the innovation standard deviations. One concern is that a diffuse prior over standard deviations can still be informative for model comparison, especially for σ_ν, which controls how much the geometric trend g_t can move. We therefore repeat the posterior estimation and model-comparison exercise under several alternatives: a weak lognormal prior, an MLE-centered lognormal prior, a prior that puts high mass on larger σ_ν, a fat-tailed half-t prior, and a persistence-friendly autoregressive specification that puts high prior mass on persistent b_t and g_t.[7]

Appendix Table B1 reports posterior summaries and final model-comparison values. The prior choices move the posterior distribution for σ_ν substantially, and the autoregressive case also pushes the posterior distribution for trend persistence toward high values. The final posterior probability weight of model $\mathcal{A}$ remains at least 0.993 in every case. Figure 12 shows that the smoothed paths and posterior weight path are very similar across priors. Thus, the result is not driven by a prior that forces the geometric trend to be too smooth or too volatile.

[7] Additional details of the priors explored are provided in Appendix B.3.

Figure 12: Prior Robustness: Smoothed States and Model Evaluation



Notes: The panels compare the smoothed trend components and posterior model weight across alternative prior specifications. The b_t and g_t panels report posterior medians. The posterior-probability panel reports the pooled posterior-predictive weight of model $\mathcal{A}$ at a 10-year forecast horizon for each prior specification, averaging predictive densities across posterior draws before accumulating log scores. When computing the posterior probability weight, the initial prior of model $\mathcal{G}$ is 0.9 and $\gamma = 0.25$.

6.3 Autoregressive Specification for Trend Growth

We next extend the baseline stochastic trend specifications by allowing the latent growth components to follow stationary autoregressive processes. This formulation nests the random-walk trend models as a limiting case and permits explicit estimation of persistence and long-run trend growth. We now specify model $\mathcal{A}$ (additive growth) as

$$\begin{aligned} A_t &= A_{t-1} + b_t + \sigma_a \epsilon_t^a \\ b_t &= (1-\rho_b)b + \rho_b b_{t-1} + \sigma_u u_t \end{aligned}$$

where $\rho_b \in (0,1)$ governs persistence, b is the long-run mean growth rate, and u_t and ϵ_t^a are independent standard normal shocks, as before. Similarly, we now specify model $\mathcal{G}$ (geometric growth) as

$$\begin{aligned} A_t &= A_{t-1}\left(1+g_t\right)\exp\left(\sigma_g \epsilon_t^g\right) \\ g_t &= (1-\rho_g)g + \rho_g g_{t-1} + \sigma_\nu \nu_t \end{aligned}$$

where $\rho_g \in (0,1)$ governs persistence, g is the long-run mean growth rate, and ν_t and ϵ_t^g are independent standard normal shocks.

For each model, we specify priors, combine them with the Kalman filter likelihood, and sample the resulting posterior distribution using MCMC. The persistence parameters ρ_b and ρ_g follow Beta distributions on $(0,1)$, the innovation variances σ_u^2, σ_a^2, σ_ν^2, and σ_g^2 follow inverse-gamma priors, and the long-run mean growth parameters b and g are assigned diffuse uniform priors.

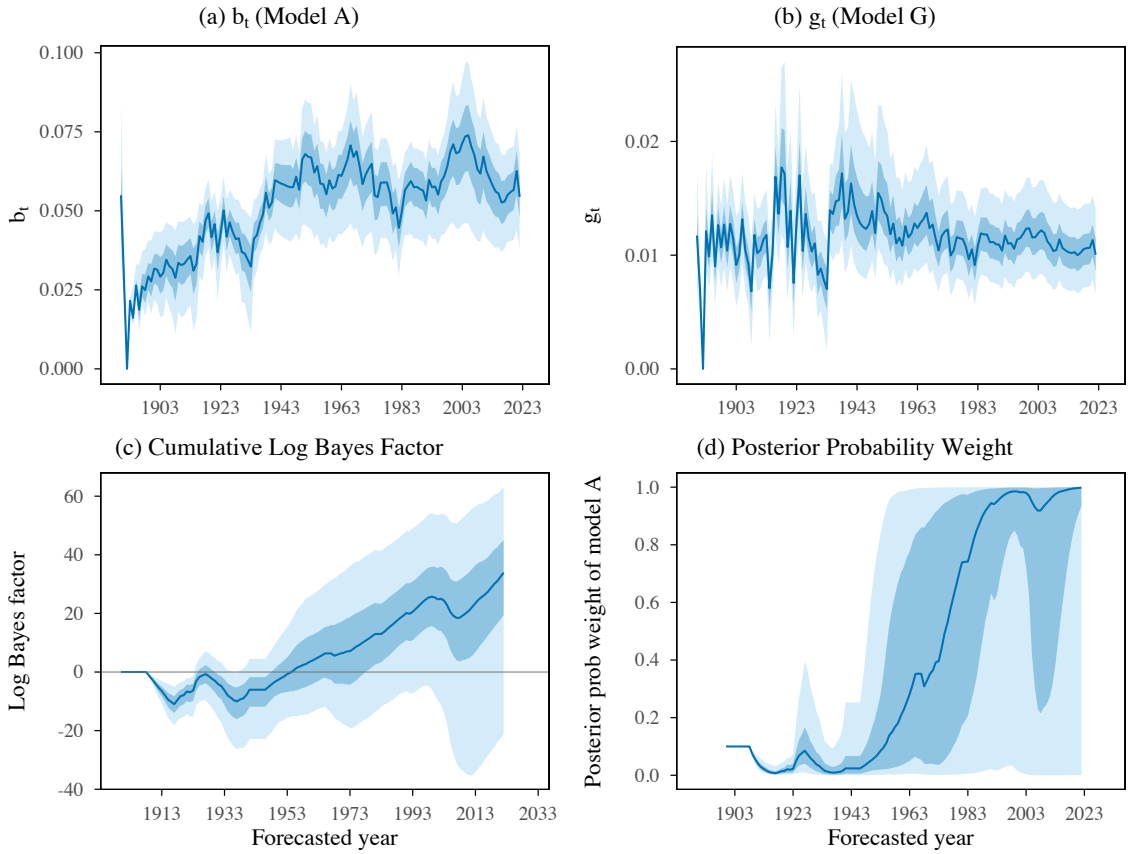
Table 3 presents the mode and standard deviation of the posterior distributions. Both models $\mathcal{A}$ and $\mathcal{G}$ exhibit persistence in their mean growth parameters, more-so for model $\mathcal{A}$ (around 0.94 for model $\mathcal{A}$ and around 0.7 for model $\mathcal{G}$). Consequently, the filtered path of b_t, shown in panel (a) of Figure 13, looks very similar to our baseline random walk specification. In addition, panel (b) shows that the filtered path of g_t follows the same broad pattern as our baseline random walk specification (high in the 1940s and 1950s, before generally gradually declining through to the end of the sample), though with more noise. Panels (c) and (d) show that the conclusions regarding the Bayesian preferred model are similar to those of our baseline specification, with the posterior model weight of model $\mathcal{A} \approx 1$ towards the end of the sample.

Table 3: Bayesian Parameter Estimates, Autoregressive Model (US, 1880-2022)

	Model $\mathcal{A}$			
	ρ_b	σ_u	σ_a	b
Mode	0.9442	0.0064	0.0677	0.0545
Standard Deviation	(0.2031)	(0.0024)	(0.0046)	(0.0219)
	Model $\mathcal{G}$			
	ρ_g	σ_ν	σ_g	g
Mode	0.7350	0.0052	0.0313	0.0113
Standard Deviation	(0.2948)	(0.0013)	(0.0021)	(0.0046)

Notes: This table provides the mode and standard deviation of the posterior distributions of the parameters governing the standard deviations of the $\mathcal{A}$ and $\mathcal{G}$ models with the autoregressive specification.

Figure 13: Autoregressive Specification



Notes: Panels (a) and (b) show the filtered values for b_t and g_t for 1,000 draws of the parameters from the posterior distributions. Panel (c) shows the cumulative log Bayes factor and panel (d) shows the posterior probability weight of model $\mathcal{A}$. The dark blue areas show the 25th and 75th percentiles of the resulting distribution and the light blue areas show the corresponding 5th and 95th percentiles. The initial prior on model $\mathcal{G}$ is 0.9, and $\gamma = 0.25$.

6.4 Piecewise Regime Breaks

In this section, we explore the robustness of our results using estimations with structural breaks. We consider an estimation treating b and g as parameters to be estimated. In Appendix B, we consider a separate estimation with structural breaks in the volatility of the shocks.

6.4.1 Piecewise b and g

We estimate versions of the models in which the drift (b_t) in the additive model or the growth rate (g_t) in the geometric model follows a piecewise constant process, with a fixed number of regimes and unknown break dates. We now specify model $\mathcal{A}$ as

$$A_t = A_{t-1} + b_t + \sigma_a \varepsilon_t^a, \quad b_t = b^{(k)} \text{ for } t \in \mathcal{R}_k,$$

where ε_t^a is a standard normal shock, and model $\mathcal{G}$ as

$$A_t = A_{t-1}(1 + g_t)\exp(\sigma_g \varepsilon_t^g), \quad g_t = g^{(k)} \text{ for } t \in \mathcal{R}_k,$$

where ε_t^g is a standard normal shock. In both of these specifications, $\mathcal{R}_k$ denotes the k^{th} regime, and the parameters $\{b^{(k)}\}$ or $\{g^{(k)}\}$ are constant within regimes but allowed to differ across them.

We consider the following specification: the number of regimes is fixed at $K_A = 3$ for Model $\mathcal{A}$ and $K_G = 3$ for Model $\mathcal{G}$ (i.e. two breakpoints per model); the innovation variances σ_a and σ_g are constant across time; regime break dates $\{t_1, \ldots, t_{K-1}\}$ are unknown and estimated; and regime lengths have a minimum length of 20 periods.

Estimation procedure. The posterior over parameters and breakpoints is explored using a Metropolis-within-Gibbs sampler, which is largely standard. It has the following steps:

1. **Initialize** the parameters $\theta = \{b^{(1)}, \ldots, b^{(K)}, \sigma_a\}$ and breakpoints.
2. **Gibbs Step:** Alternate between:
   (a) **Parameter Step:** Propose a new θ' using a random-walk Metropolis step:

$$\theta' = \theta + \kappa \cdot \text{chol}(H^{-1}) \cdot \eta, \quad \eta \sim \mathcal{N}(0, I)$$

where H is the negative Hessian from the initial maximum likelihood estimation in which the regimes are equally spaced over the sample.

(b) **Breakpoint Step:** Propose new breakpoints using a local random shift mechanism that maintains the minimum separation constraint. Accept or reject the new breakpoints using a Metropolis step based on the change in log-posterior.

We use uninformative priors on the breakpoints and impose independent log-normal priors on the growth parameters and the standard deviations. An adaptive tuning procedure targets an acceptance rate of around 30%. We retain 1,000,000 posterior draws after a burn-in of 100,000 iterations.

6.4.2 Estimation Results

Figure 14 plots the posterior estimates of b and g with structural breaks. The estimates for b are consistent with those of the random walk filtered values for b_t under the MLE benchmark estimates shown above, with a notable break in the late 1930s, after which it remains roughly stable. The estimates for g are also roughly consistent with the estimated path of g_t under the MLE benchmark estimates shown above, with a step up in the estimate value around the 1940s, and a step down around the 1960s. Panels (c) and (d) show that this specification again assigns posterior probability weight near one to model $\mathcal{A}$ by the end of the sample.

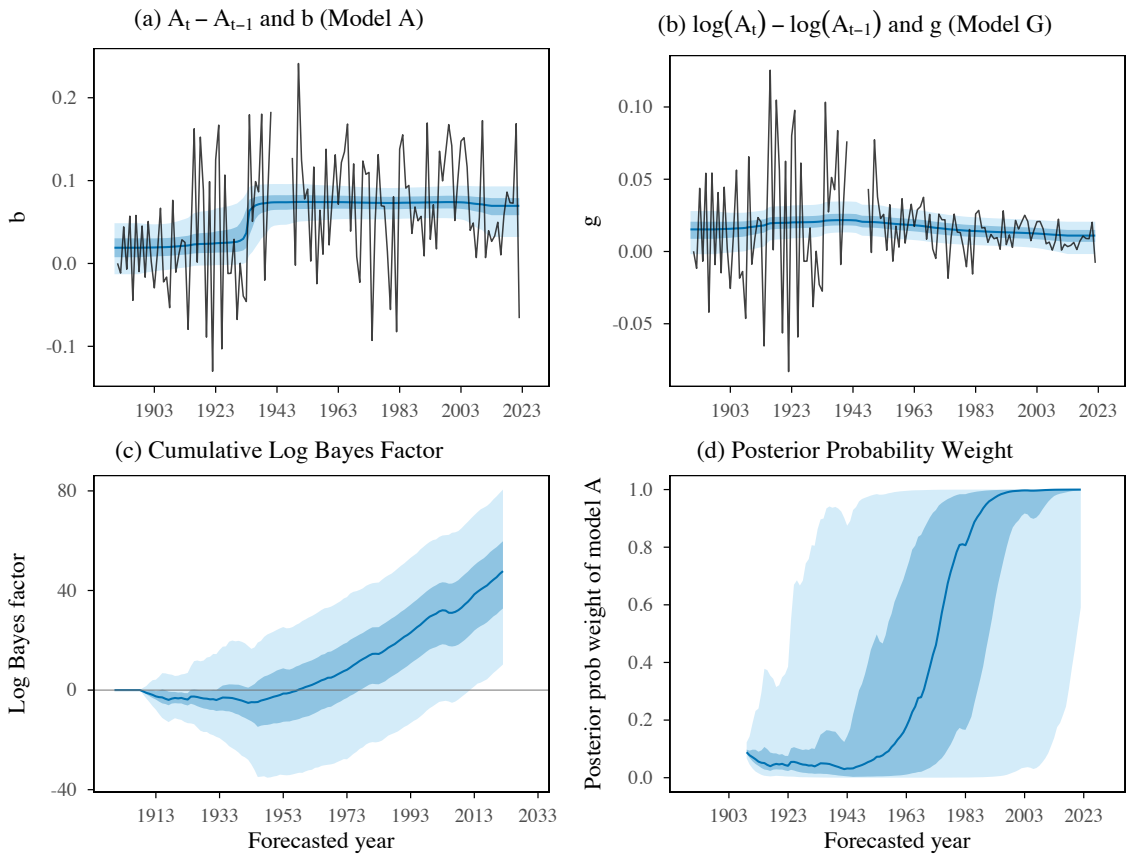
In Appendix B.5, we additionally consider the case where the number of regimes in the $\mathcal{A}$ model for b is set to two (i.e. $K_A = 2$, or one break in b), and compare that to the case where there are three regimes for g for Model $\mathcal{G}$ (two breaks in g). This choice allows for more flexibility for the geometric model compared to the additive model. As shown in the Appendix, under this specification, we find that the posterior model weight on model $\mathcal{A} \approx 1$ by the end of the sample.

7 Conclusion

U.S. TFP has been growing linearly over the past 90 years, and the additive model outperforms the exponential model in our U.S. model-comparison exercises.

The additive model, unlike the exponential one, provides useful long-term forecasts of TFP. Supporting evidence from professional forecasts and international TFP series

Figure 14: Piecewise b and g Specification



Notes: Both the $\mathcal{A}$ and $\mathcal{G}$ models allow for two breaks in the trend growth parameters b and g. In panels (a) and (b), the data is plotted in black and the median estimate of b and g over time is plotted as the blue line. Panel (c) shows the cumulative log Bayes factor and panel (d) shows the posterior probability weight of model $\mathcal{A}$. The dark blue areas show the 25th and 75th percentiles of the resulting distribution and the light blue areas show the corresponding 5th and 95th percentiles. The initial prior on model $\mathcal{G}$ is 0.9, and $\gamma = 0.25$.

points in the same direction.

The additive growth model explains the observed TFP slowdown as a consequence of model misspecification. We should not have expected growth *rates* to be constant in the first place. Additive TFP growth predicts *increasing* increments of labor productivity, wages, and GDP per capita thanks to capital accumulation.

To conclude, we mention three important points. Firstly, additive growth has implications for macroeconomic, industry and firms dynamics. Wealth effects of TFP news are weaker and long run consumption risk is lower. Additive growth also matters for the optimal mitigation of climate change, since real discount rates are low and

future generations will not be much richer than the current one.

Secondly, at the firm level, Lenzu et al. (2023) finds that productivity grows linearly with the age of a firm. More importantly, the study of industries and firms can shed light on *why* growth is additive and *where* knowledge transfers take place. At a more theoretical level, additive growth suggests that new ideas add to our stock of knowledge but do not multiply it. This observation does not rule out positive spillover from R&D across firms, and especially across vintages of firms.

Thirdly, the additive model allows us to make long-term forecasts. The work of Fernandez-Villaverde et al. (2026) provides an interesting comparison. Assuming that the productivity transfer function between the U.S. and China follows the pattern previously observed in fast-growing Asia economies, they predict that China's income per capita should reach about 40% of the U.S. level by 2100. It is a matter of simple arithmetic to create such a forecast in our model. By the end of 2019, Chinese GDP per capita was around 26% of US GDPPC. We translate this figure into a relative TFP number (about 0.4), then use our estimates of TFP increments from the Kalman filter. The increments are remarkably similar in both countries. Using these estimates and a 80 years forecast horizon, our model predicts that Chinese TFP will be around 2/3 of US TFP by 2100, with a corresponding relative GDPPC of slightly more than 1/2. Despite the different approaches, the predicted values are remarkably similar.

References

Akcigit, U. and S. T. Ates (2023): "What Happened to US Business Dynamism?" *Journal of Political Economy*, 131, 2059–2124.

Bergeaud, A., G. Cette, and R. Lecat (2016): "Productivity Trends in Advanced Countries between 1890 and 2012," *Review of Income and Wealth*, 62, 420–444.

Bloom, N., C. I. Jones, J. Van Reenen, and M. Webb (2020): "Are Ideas Getting Harder to Find?" *American Economic Review*, 110, 1104–44.

Bouscasse, P., E. Nakamura, and J. Steinsson (2025): "When Did Growth Begin? New Estimates of Productivity Growth in England from 1250 to 1870," *The Quarterly Journal of Economics*, 140, 835–888.

Comin, D., W. Easterly, and E. Gong (2010): "Was the Wealth of Nations Determined in 1000 bc?" *American Economic Journal: Macroeconomics*, 2, 65–97.

David, P. A. (1990): "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox," *The American Economic Review*, 80, 355–361.

Del Negro, M., R. B. Hasegawa, and F. Schorfheide (2016): "Dynamic prediction pools: An investigation of financial frictions and forecasting performance," *Journal of Econometrics*, 192, 391–405, innovations in Multiple Time Series Analysis.

DeLong, J. B. (2022): *Slouching Towards Utopia: An Economic History of the Twentieth Century*, New York, NY: Basic Books.

Edge, R., T. Laubach, and J. C. Williams (2007): "Learning and shifts in long term productivity growth," *Journal of Monetary Economics*, 54, 2421–2438.

Fernandez-Villaverde, J., L. Ohanian, and W. Yao (2026): "The Neoclassical Growth of China," Working Paper.

Field, A. J. (2003): "The Most Technologically Progressive Decade of the Century," *American Economic Review*, 93, 1399–1413.

Friedman, M. (1953): *Essays in Positive Economics*, University of Chicago Press.

Gordon, R. J. (2016): *The Rise and Fall of American Growth*, Princeton University Press.

Guzey, A., E. Rischel, Applied Divinity Studies, Anonymous, and Anonymous (2021): "Issues with Bloom et al's "Are Ideas Getting Harder to Find?" and why total factor productivity should never be used as a measure of innovation," https://guzey.com/economics/bloom/.

Jones, B. F. (2009): "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" *The Review of Economic Studies*, 76, 283–317.

Jones, C. I. (1995): "RD-Based Models of Economic Growth," *Journal of Political Economy*, 103, 759–784.

Lenzu, S., T. Philippon, and J. Tielens (2023): "Additive Firm Dynamics," NYU WP.

Philippon, T. (2022): "Additive Growth," Working Paper.

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

——— (1957): "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics*, 39, 312–320.

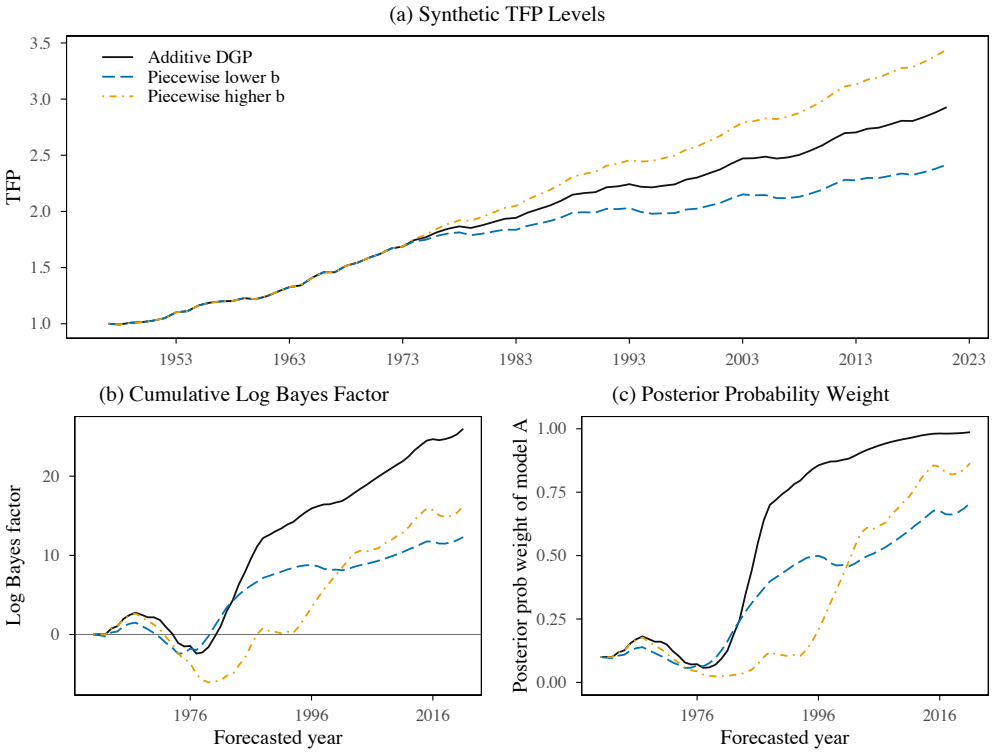
Appendix

A US Data

A.1 Exercise with Simulated Additive Data

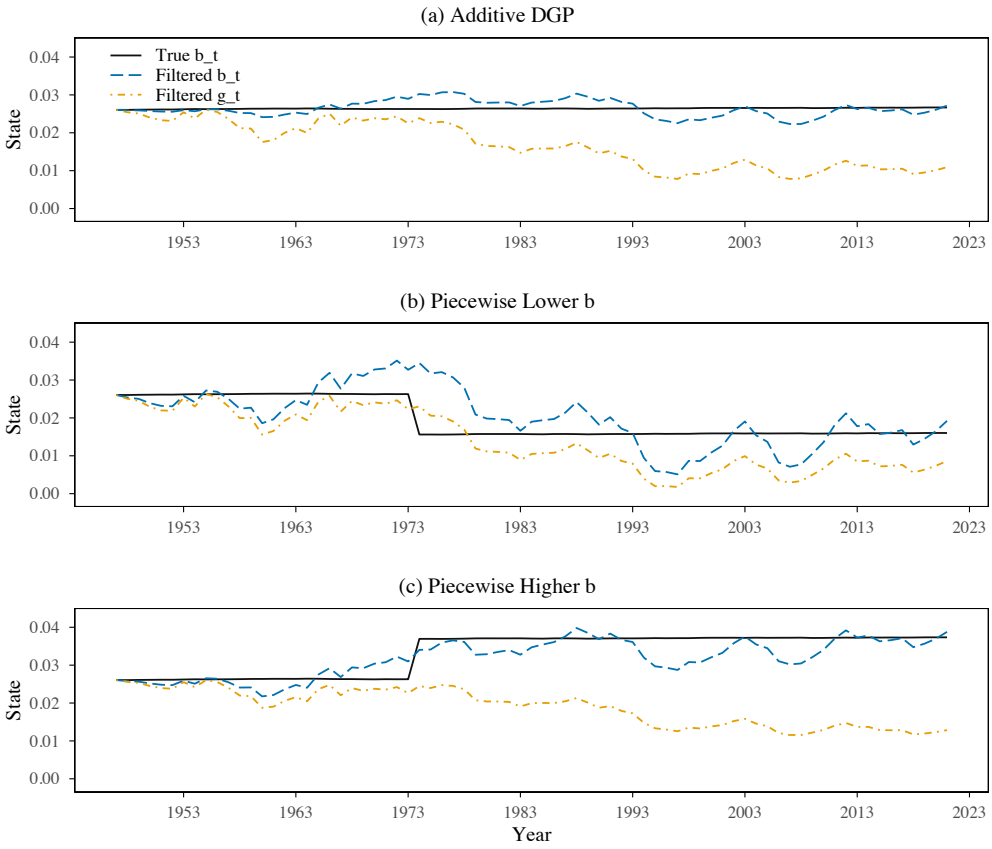
Figure 7 illustrated the comparison of the U.S. TFP series in BCL and Fernald, with and without a labor-quality adjustment. Here, we use simulated data to check whether the lower posterior weight in the labor-quality-adjusted Fernald series can arise when the true data-generating process is additive. We simulate annual TFP series of the same length as the Fernald sample using the BCL common-sample estimates of the additive model. In the first simulation, b_t follows the estimated additive random-walk

Figure A1: Simulated Additive Data and Model Fit



Notes: The figure reports synthetic TFP paths and posterior-predictive model comparisons at the 10-year horizon. The synthetic series have the same length as the Fernald sample and are generated from the additive model using the BCL common-sample estimates. The piecewise cases impose a one-time shift in b_t in 1974. In the re-estimation, both filters are initialized at the true initial trend, and the priors impose $\sigma_u \geq 0.001$ and $\sigma_\nu \geq 0.001$.

Figure A2: Filtered Trends in Synthetic Additive Data



Notes: The black line is the true simulated b_t. The blue line is the filtered b_t from model $\mathcal{A}$, and the orange line is the filtered g_t from model $\mathcal{G}$. The filters are initialized at the true initial trend. Filtered paths use the posterior-mode parameter vector from the re-estimation with $\sigma_u \geq 0.001$ and $\sigma_\nu \geq 0.001$.

law with no deterministic break. In the second and third simulations, we introduce an unanticipated one-time reduction or increase in b_t in 1974, with the shift calibrated to the decline in average BCL level increments around the postwar slowdown. We then re-estimate the additive and geometric models on each simulated series and repeat the 10-year posterior-predictive model comparison.

Figure A1 reports the simulated TFP paths and the resulting model-comparison statistics. The pure additive simulation gives strong posterior-predictive support to model $\mathcal{A}$. A downward break in b_t, however, lowers the posterior probability weight, even though the data are generated by an additive process. This exercise therefore supports the interpretation of Figure 7, or that a flatter postwar TFP series, or one with a discrete slowdown in additive increments, weakens the relative contrast

between additive and geometric forecasts without implying that the geometric model is the true data-generating process.

Figure A2 shows the true simulated b_t alongside the filtered b_t and g_t paths from the re-estimated models. The filtered additive trend tracks the level and break in the true b_t. The geometric trend instead moves down over time, especially after the slowdown, because it tries to represent additive level increments as a percentage growth rate.

A.2 Survey of Professional Forecasters

The data from the Survey of Professional Forecasters is summarized in the following table.

Variable Name	Description	Date Availability
PROD10	Quarterly forecasts of annual-average productivity growth rates, where rates are annualized and productivity is measured by output per hour, from the survey date to 10 years later (40 quarters total). Obtained from Survey of Professional Forecasters.	1992:Q1 to 2022:Q3

A.3 Council of Economic Advisors

This section summarizes the projection data obtained from the Council of Economic Advisors (CEA).

The CEA projections for output per hour growth are taken from the annual Economic Report of the President. These forecasts are consistently available since 1983, with the exception of 1995 and 2013. Forecasts are also available in the 1970 and 1971 Reports, but with indefinite time horizons. While the values for 1970 and 1971 must be inferred from the text of the Report itself, all other values are gathered from tables in the report. The numbers of these tables in each report are listed in the supplemental data. These tables are titled the following:

- 1983: "Projections of Economic Goals,"
- 1984 to 1991: "Administrative Economic Assumptions,"

- 1992 to 2005: "Accounting for Growth in Real GDP,"
- 2006 to 2009: "Supply-Side Components of Real GDP Growth,"
- 2010 to 2011: "Components of Potential Real GDP Growth,"
- 2012: "Components of Actual and Potential Real GDP Growth,"
- 2014 to 2022: "Supply-Side Components of Actual and Potential Real GDP Growth."

Time horizons of the forecasts fluctuate over the course of the data. Between 1983 and 1994, these forecasts have a six-year time horizon, including the year of the report. From 1996 to 2002, the forecasts' time horizons range from seven to eleven years out, before returning to a six-year forecast time horizon until 2009. Following 2009, the forecasts consistently have an eleven year time horizon, with the exception of the 2013 Report with no forecasts.

The granularity of the forecasts also varies over the reports. The 1970 and 1971 reports forecast 2.8 and 3 percent annual growth of output per hour in perpetuity. Meanwhile, between 1983 and 1991, each projected year receives its own distinct forecast value. However, after 1991, all Reports give only a single per annum forecast value, always over a period that includes at least a year–if not more–of realized values.

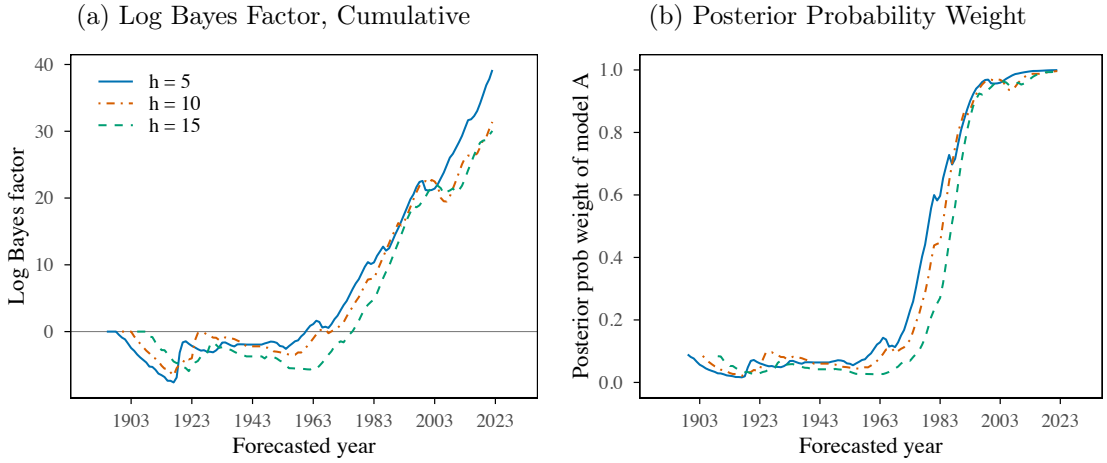
B Additional Results using U.S. Data

B.1 Point Forecast Evaluation

This section complements the density-based model comparison in the main text by evaluating the point forecast performance of the $\mathcal{A}$ and $\mathcal{G}$ models. The goal is to isolate differences in conditional mean dynamics.

As discussed in Section 4, the predictive density comparison jointly reflects differences in both the mean and variance implied by the two models. Here, we conduct a point forecast evaluation based on mean squared error (MSE). For each model, we construct h-step-ahead point forecasts using the conditional mean implied by the estimated state-space representation. Under the additive model, the h-step-ahead forecast for the level of the series is given by $A_{t+h} = A_t + h\, b_t$, where b_t denotes the filtered drift. Under the geometric model, the corresponding level forecast is

Figure B3: Posterior Model Probability, $h = 5, 10, 15$ Year Forecast Horizons



Notes: When computing the posterior probability weight, the initial prior of model $\mathcal{G}$ is 0.9.

$A_{t+h} = A_t \exp(h\, g_t)$, where g_t is the filtered growth rate. In both cases, forecasts are evaluated against realized levels, and forecast errors are computed in levels.

Using the point forecasts implied by each model's conditional mean, we compute mean squared forecast errors for an 10-year-ahead horizon. The $\mathcal{A}$ model yields an MSE of 0.068, compared with 0.158 for the $\mathcal{G}$ model. In root mean squared error terms, the $\mathcal{A}$ model has a RMSE of 0.26, while the $\mathcal{G}$ model has a RMSE of 0.40. The resulting MSE ratio of about 0.43 indicates that the additive model reduces mean squared forecast error by more than half relative to the geometric alternative.

B.2 Other Forecast Horizons

The baseline forecast horizon used to evaluate the $\mathcal{A}$ and $\mathcal{G}$ models in the Bayesian model evaluation is 10 years. This section instead presents the Bayesian model evaluation under 5 and 15 year forecast horizons, alongside the baseline 10 year forecast horizon for comparison. Figure B3 shows the cumulative log Bayes factor and the posterior probability weight. The results are consistent with the 10-year forecast evaluation: by the end of the sample, the $\mathcal{A}$ model is preferred with a posterior weight ≈ 1. The model evaluation also fluctuates less around the prior at shorter horizons before the 1970s.

Table B1: Prior Robustness

Prior	Model	Final log BF	Final $P(\mathcal{A})$	Mode σ_ν	P90 σ_ν	Mode ρ_b	P90 ρ_b	Mode ρ_g	P90 ρ_g
Uniform sigma	RW trend	31.52	0.997	0.0006	0.0028	–	–	–	–
Weak lognormal	RW trend	30.31	0.995	0.0004	0.0021	–	–	–	–
Lognormal around MLE	RW trend	30.98	0.996	0.0002	0.0011	–	–	–	–
High sigma_nu stress	RW trend	35.24	0.999	0.0021	0.0045	–	–	–	–
Fat-tailed half-t	RW trend	31.88	0.997	0.0006	0.0030	–	–	–	–
Persistence-friendly AR	AR trend	37.60	0.999	0.0012	0.0052	0.967	0.983	0.880	0.959

Notes: This table reports posterior summaries for selected prior specifications and the final model-comparison values at the 10-year forecast horizon. The posterior probability weight is the posterior probability weight of model $\mathcal{A}$. When computing the posterior probability weight, the initial prior of model $\mathcal{G}$ is 0.9 and $\gamma = 0.25$.

B.3 Prior Robustness

Table B1 reports posterior summaries and final model-comparison values for the prior-robustness exercise discussed in Section 6.2. For the random-walk trend specifications, the sampler works with the log standard deviations, ordered as (σ_u, σ_a) for model $\mathcal{A}$ and (σ_ν, σ_g) for model $\mathcal{G}$. The uniform case puts independent uniform priors on each standard deviation over $[\epsilon, 1]$. The weak lognormal case sets $\log \sigma_j \sim N(\log m_j, 1.5^2)$, with $m = (0.0025, 0.05)$ in the same parameter ordering. The MLE-centered lognormal case sets $\log \sigma_j \sim N(\log \hat{\sigma}_j^{MLE}, 1^2)$. The high-$\sigma_\nu$ stress test keeps the MLE-centered lognormal prior for model $\mathcal{A}$, but for model $\mathcal{G}$ sets $\log \sigma_\nu \sim N(\log 0.005, 0.5^2)$ and $\log \sigma_g \sim N(\log \hat{\sigma}_g^{MLE}, 1^2)$. The fat-tailed case uses independent half-t priors with 3 degrees of freedom, scales $(0.01, 0.10)$, and support $[\epsilon, 1]$. The persistence-friendly AR case replaces the random-walk trend laws with AR(1) trend laws. It puts $\rho_b, \rho_g \sim \text{Beta}(8, 1.5)$, uses the same half-$t$ priors on the two innovation standard deviations in each model, and puts a $N(0.02, 0.10^2)$ prior on each steady-state trend mean.

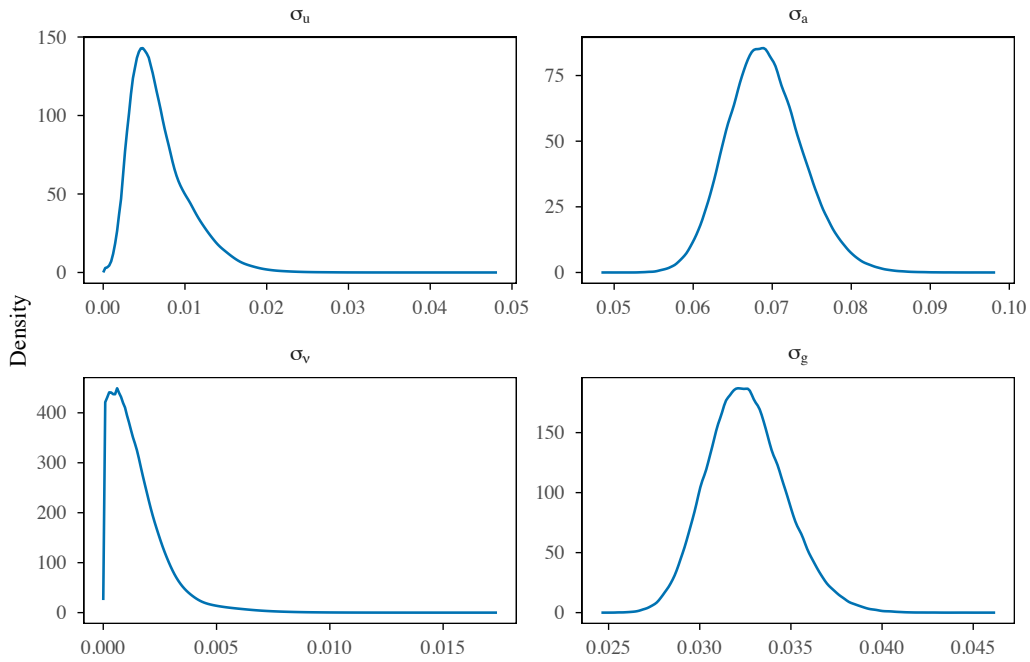
B.4 Bayesian Estimation of the Parameters

The algorithm we use is standard and is described next.

Metropolis-Hastings algorithm. Given current parameter vector θ, a new proposal θ' is drawn using a random walk based on a Student-t proposal with 12 degrees of freedom:

$$\theta' = \theta + \kappa \cdot \text{chol}\left(H^{-1}\right) \cdot \varepsilon, \quad \varepsilon \sim t_{12}\left(0, I\right),$$

Figure B4: Posterior Density of Estimated Parameters



Notes: This figure shows the posterior distributions of the parameters governing the standard deviations of the $\mathcal{A}$ and $\mathcal{G}$ models.

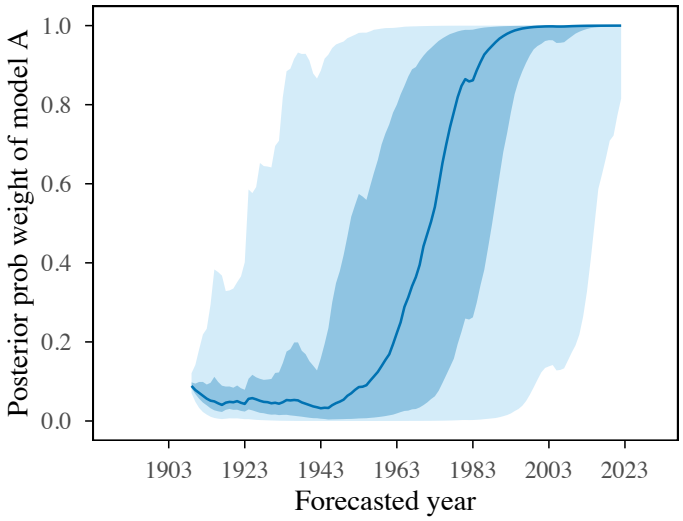
where H is the Hessian from the log-likelihood function evaluated at the maximum likelihood estimate, and κ is a tuning constant for step size. The acceptance probability is computed as:

$$\alpha = \min\left(1, \exp\left[\log p\left(y \mid \theta'\right) + \log \pi\left(\theta'\right) - \log p\left(y \mid \theta\right) - \log \pi\left(\theta\right)\right]\right),$$

where $p\left(y \mid \theta\right)$ is the likelihood (evaluated using the Kalman filter), and $\pi\left(\theta\right)$ is the prior. During the burn-in period, the scaling factor κ is updated every 1000 iterations to target an acceptance rate of 30%.

Figure B4 plots the posterior densities of the estimated parameters, showing they are single peaked and their modal values are close to those of the MLE benchmark estimation.

Figure B5: Posterior Prob Weight of $\mathcal{A}$ Model, Piecewise Model, $K_{\mathcal{A}} = 2$, $K_{\mathcal{G}} = 3$.



Notes: The median is plotted as the blue line. The dark blue areas show the 25th and 75th percentiles of the resulting distribution and the light blue areas show the corresponding 5th and 95th percentiles. The initial prior on model $\mathcal{G}$ is 0.9, and $\gamma = 0.25$.

B.5 Estimations with Structural Breaks

B.5.1 Regime Breaks in Trend Growth

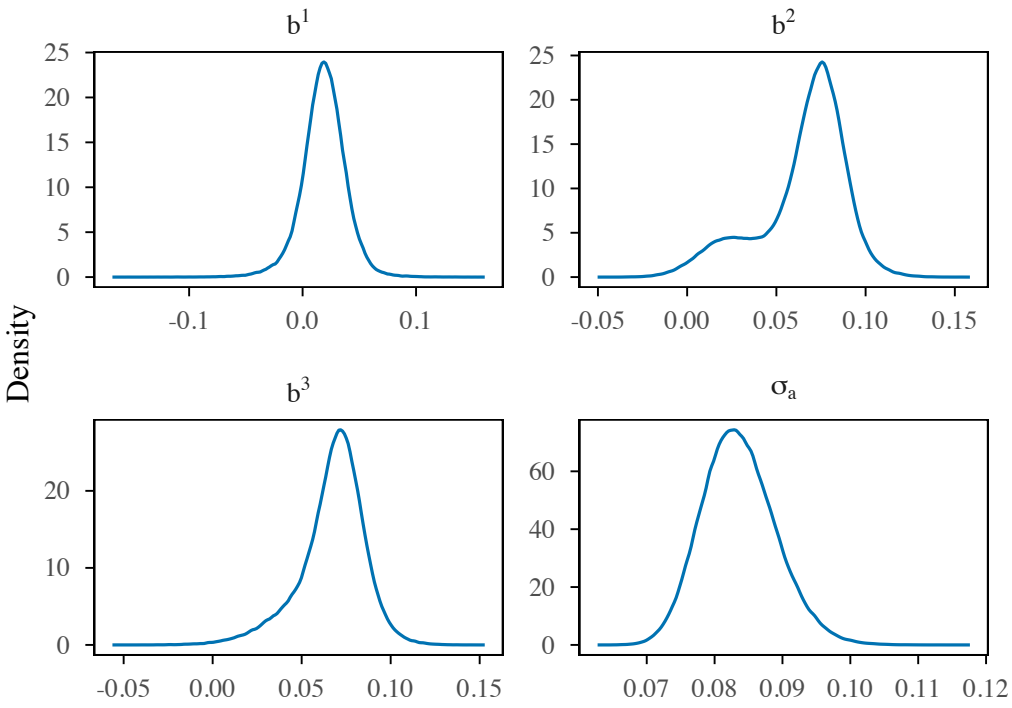
This section contains additional results for the estimation of the model with regime breaks in the growth parameters b and g. Figure B5 shows the posterior probability weight of the $\mathcal{A}$ model under the piecewise break specification, allowing for one break in b in the $\mathcal{A}$ model and two breaks in g in the $\mathcal{G}$ model. Consistent with the results shown in the main text, under these specifications we find that the posterior model weight of model $\mathcal{A} \approx 1$ towards the end of the sample.

B.5.2 Posterior Estimates of Baseline Model with Regime Breaks

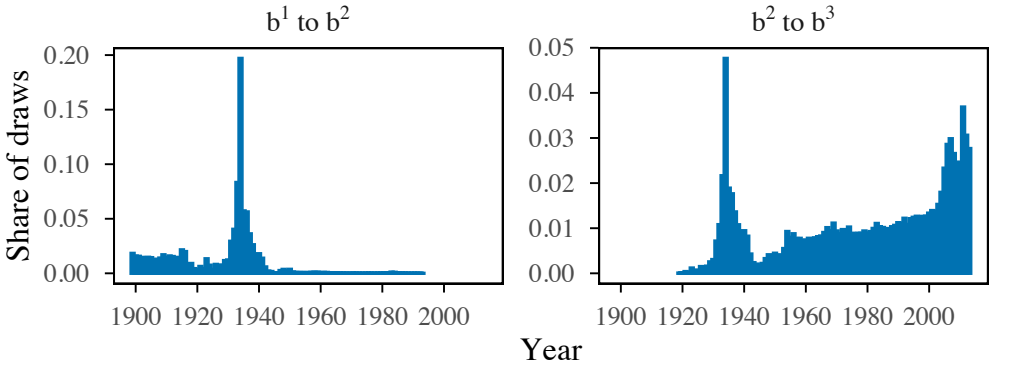
Figures B6 and B7 show the posterior distributions of the parameters under the piecewise specifications for b and g in which we allow for three regimes in either model. Panel (a) shows the estimate values of b and g across the three regimes, while Panel (b) shows the posterior distributions of the breakpoint dates for the regime breaks. For model $\mathcal{A}$, there is a clear increase in the estimate of b between the first and second regimes around the 1930s, with less evidence of a structural break between the second and third regimes (reflected in a wide posterior of the break date between

Figure B6: Posterior Densities, Model $\mathcal{A}$, Model with Regime Changes

(a) Estimated Parameters



(b) Estimated Break Dates

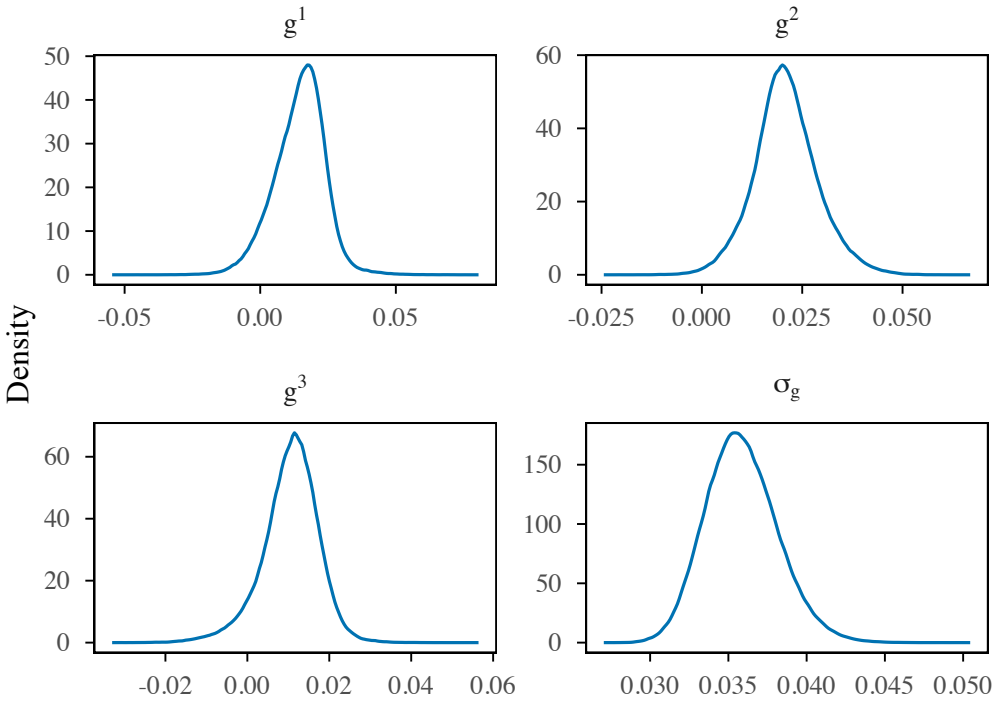


Notes: This figure shows the posterior distributions for the parameters of the $\mathcal{A}$ model with structural breaks in b.

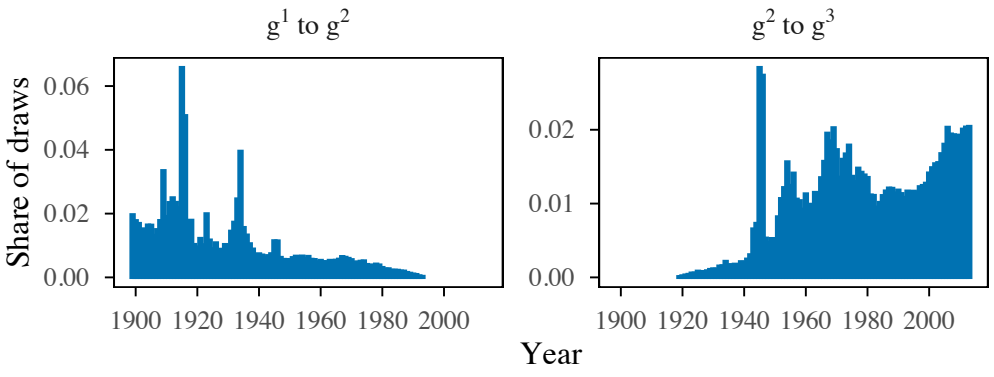
b^2 and b^3). Model $\mathcal{G}$ shows more evidence of a step-down in geometric growth around the 1950s to 1960s.

Figure B7: Posterior Densities, Model $\mathcal{G}$, Model with Regime Changes

(a) Estimated Parameters



(b) Estimated Break Dates



Notes: This figure shows the posterior distributions for the parameters of the $\mathcal{G}$ model with structural breaks in g.

B.5.3 Time-varying Variances

We next estimate the two models with time-varying stochastic volatility.

Additive model:

$$A_t = A_{t-1} + b_t + \sigma_{a,t}\,\varepsilon_t^a,$$
$$b_t = b_{t-1} + \sigma_u\,u_t,$$

where u_t and ε_t^a are independent standard normal shocks. The innovation variance in the TFP equation, $\sigma_{a,t}^2$, is allowed to change over time via a piecewise-constant structure. We assume K regimes, with regime-specific values for $\sigma_{a,t}$:

$$\sigma_{a,t} = \sigma_a^{(k)} \quad \text{for } t \in \text{Regime } k.$$

Geometric model:

$$A_t = A_{t-1}(1 + g_t)\exp(\sigma_{g,t}\,\varepsilon_t^g),$$
$$g_t = g_{t-1} + \sigma_\nu\,\nu_t,$$

where ν_t and ε_t^g are standard normal shocks. As in the additive model, the innovation standard deviation $\sigma_{g,t}$ is allowed to vary across regimes:

$$\sigma_{g,t} = \sigma_g^{(k)} \quad \text{for } t \in \text{Regime } k.$$

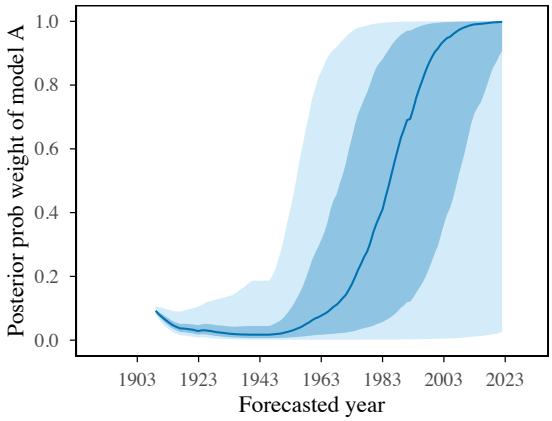
We consider the case of four regimes in the volatilities, and use a similar estimation procedure as above for the case of breaks in b and g.

Figure B8 shows the posterior probability weight of model $\mathcal{A}$ under this specification, showing as above that the weight ≈ 1 towards the end of the sample.

C International Data and Additional Results

Figure C9 plots the TFP data for all countries in the BCL dataset. Figure C10 shows the posterior probability weight for each of the countries in the BCL dataset, computed from posterior-predictive scores using country-specific posterior parameter

Figure B8: Posterior Probability Weight of Model $\mathcal{A}$, Regime Breaks in Volatility



Notes: Based on 1,000 posterior draws. Both models allow for four regimes, or three breaks, in σ_a or σ_g. The median is plotted as the blue line. The light blue shaded area shows the 5% and 95% posterior bands in light blue, the darker blue shaded area shows the 25% and 75% posterior bands.

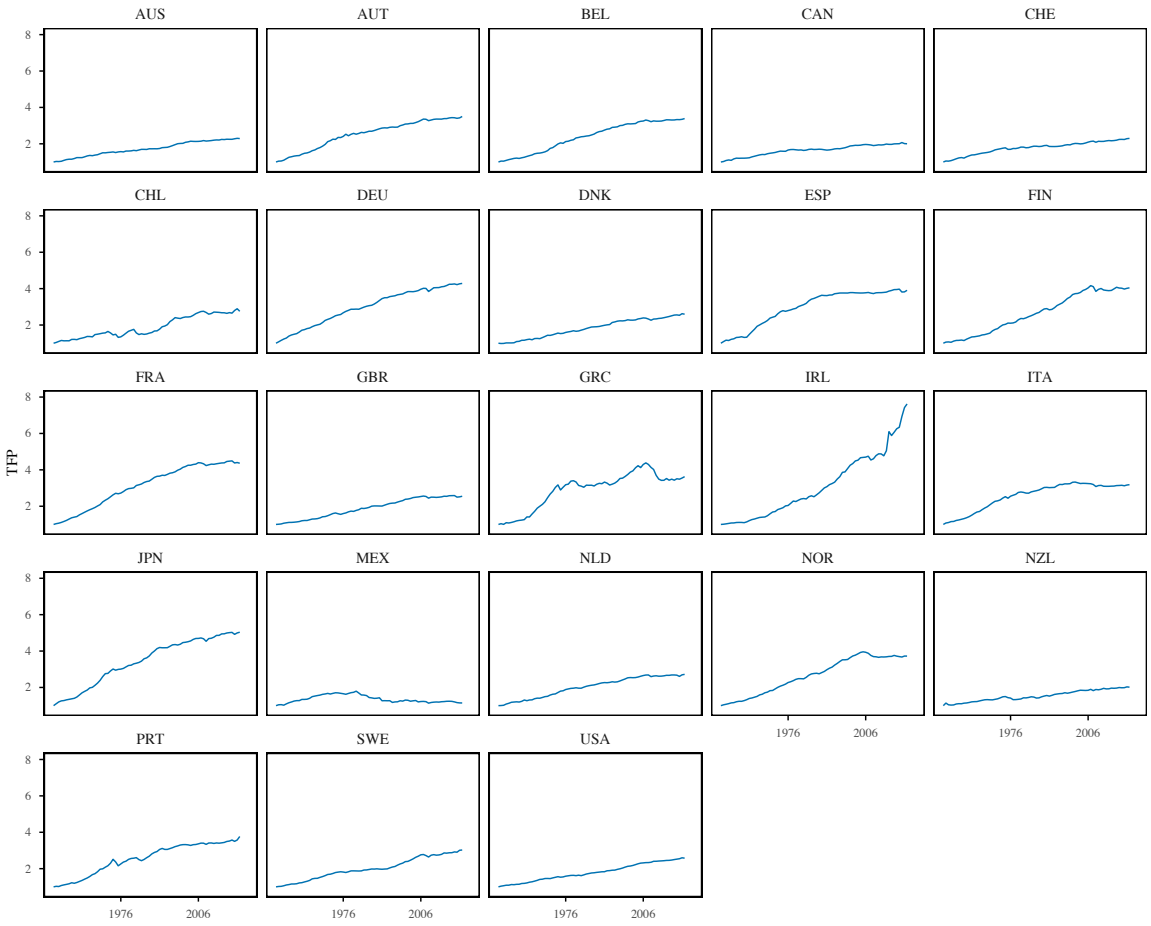
Table C2: Ratio of MSE of $\mathcal{A}$ Model to $\mathcal{G}$ Model, BCL Dataset

Country	MSE($\mathcal{A}$)/ MSE($\mathcal{G}$)	Country	MSE($\mathcal{A}$)/ MSE($\mathcal{G}$)
AUS	0.606	GRC	0.590
AUT	0.470	IRL	1.276
BEL	0.587	ITA	0.581
CAN	0.629	JPN	0.452
CHE	0.607	MEX	0.799
CHL	0.692	NLD	0.580
DEU	0.376	NOR	0.788
DNK	0.591	NZL	0.507
ESP	0.418	PRT	0.553
FIN	0.822	SWE	0.580
FRA	0.560	USA	0.588
GBR	0.642		

Notes: This table provides the ratio of the MSE of the $\mathcal{A}$ model to the MSE of the $\mathcal{G}$ model using the posterior-median filtered states from the Bayesian estimates for each country.

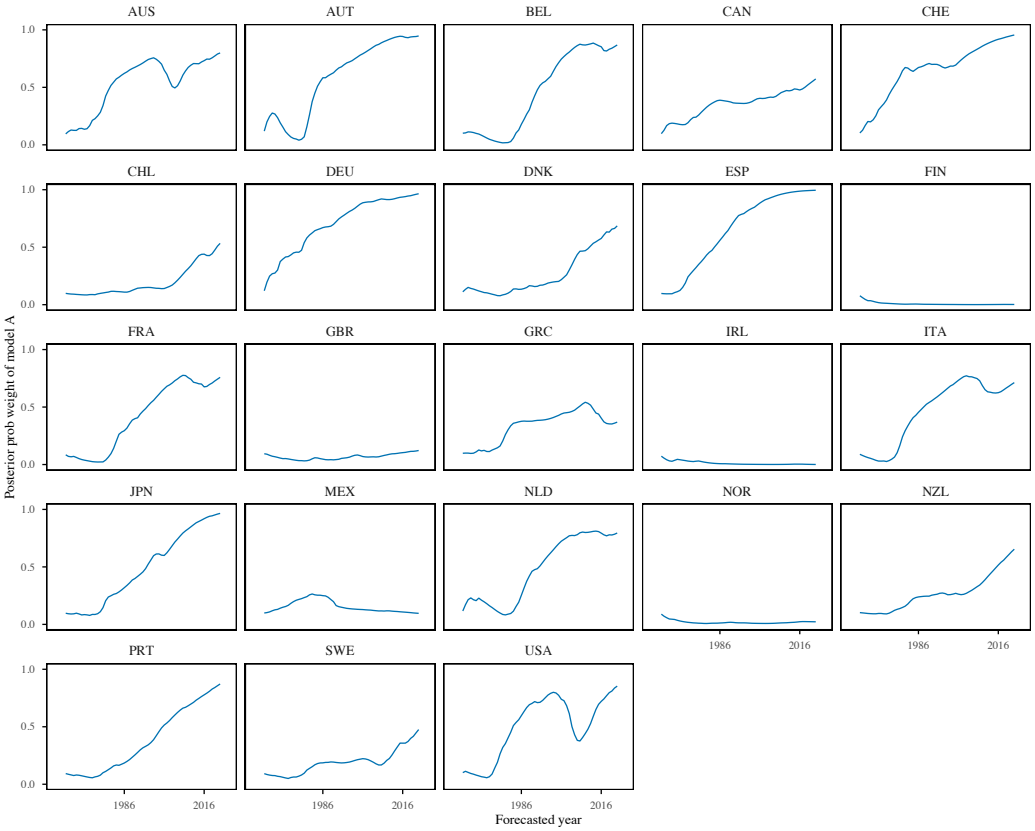
draws. Finally, Table C2 gives the ratio of a simple MSE forecast evaluation of the two models at the 10-year horizon, using posterior-median filtered states for each country. With the exception of Ireland, these results show the $\mathcal{A}$ model provides superior forecast performance across all countries.

Figure C9: TFP Levels



Notes: TFP levels for each country (1950=1). Data from Bergeaud et al. (2016).

Figure C10: Posterior Probability Weight on Model $\mathcal{A}$, BCL Sample



Notes: The figure shows the posterior probability weight for each country using posterior-predictive scores from country-specific posterior parameter draws.