Additive Growth*

Thomas Philippon†

June 2023

Abstract

Growth theory is based on the assumption of exponential total factor productivity (TFP) growth. Across countries and time periods I find that TFP growth is additive. There is no evidence that TFP increments increase with the level of TFP as predicted by the exponential model. Even starting from low priors, Bayesian estimation selects the additive model over the exponential one. The additive growth model, unlike the exponential one, provides useful long-term forecasts for TFP. For the distant past the model suggests piecewise linear evolutions with infrequent changes: the size of TFP increments increases around 1650, 1830 and 1930. For the distant future the model predicts ever increasing increments in standards of living but with falling real interest rates and growth rates that converge to zero.

*I am grateful to Tim Cogley, Chad Jones, Virgiliu Midrigan, Greg Mankiw, Olivier Blanchard, Xavier Gabaix, Xavier Jaravel, Antonin Bergeaud, Remy Lecat, David Weil, Gilbert Cette, David Romer, Callum Jones, German Gutierrez, Ben Jones, Alexey Guzey, Peter Kruse-Andersen, and Bill Easterly for their comments, and to Yad Selvakumar for outstanding research assistance.

†New York University Stern School of Business, CEPR and NBER, tphilipp@stern.nyu.edu

This paper is an empirical investigation of the stochastic process that governs total factor productivity (TFP). At least since Solow (1956) economists have assumed that TFP A_t follows an exponential process. In a deterministic setup this model (model $\mathcal{G}$ for "geometric") takes the form:

$$A_{t+\tau} = A_t e^{g\tau}, \tag{1}$$

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

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

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

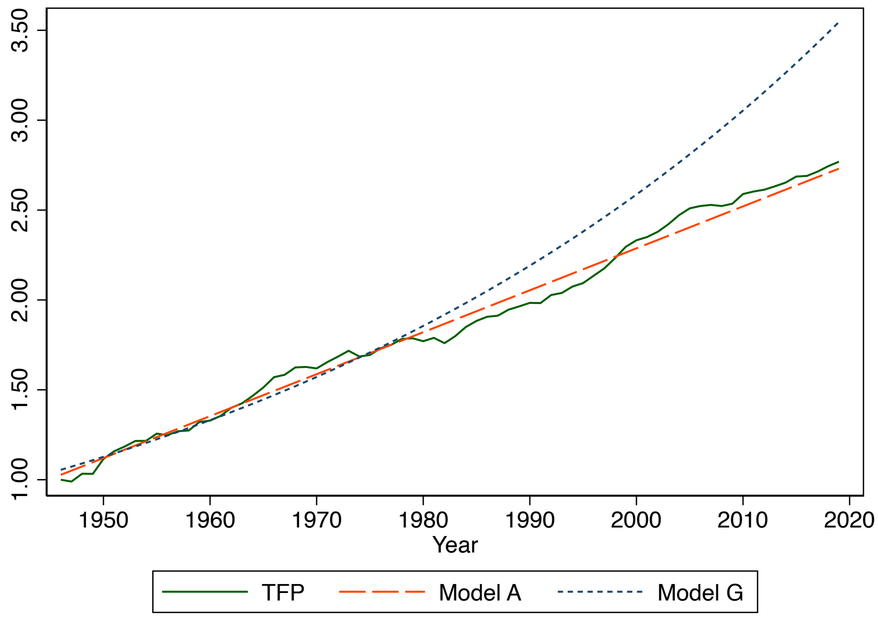
Figure 1 provides a straightforward motivation for this paper. It suggests that TFP growth has been linear in the US since at least World War II. In the figure, models $\mathcal{A}$ and $\mathcal{G}$ are estimated over the first half of the sample (1947-1983) and then used to predict the level of TFP in the second half of the sample (1984-2019). One can observe the well-known TFP slowdown "puzzle" with model $\mathcal{G}$, which simply means that actual TFP has fallen short of the exponential benchmark. By contrast there is no TFP slowdown according to model $\mathcal{A}$.

Equivalently, Figure 2 shows that the additive growth model predicts the correct path of TFP slowdown in the exponential model.

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. With the normalization 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 rate is half of what it used to be. After 60 years, it is around one percent, in line with the data. 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. The point that TFP is additive therefore does not depend on the (possibly noisy) measure of the capital stock.

The goal of this paper is to formally test the proposition that TFP growth is additive, across many countries and time periods, and then to draw the implications for economic

Figure 1: US Post War TFP



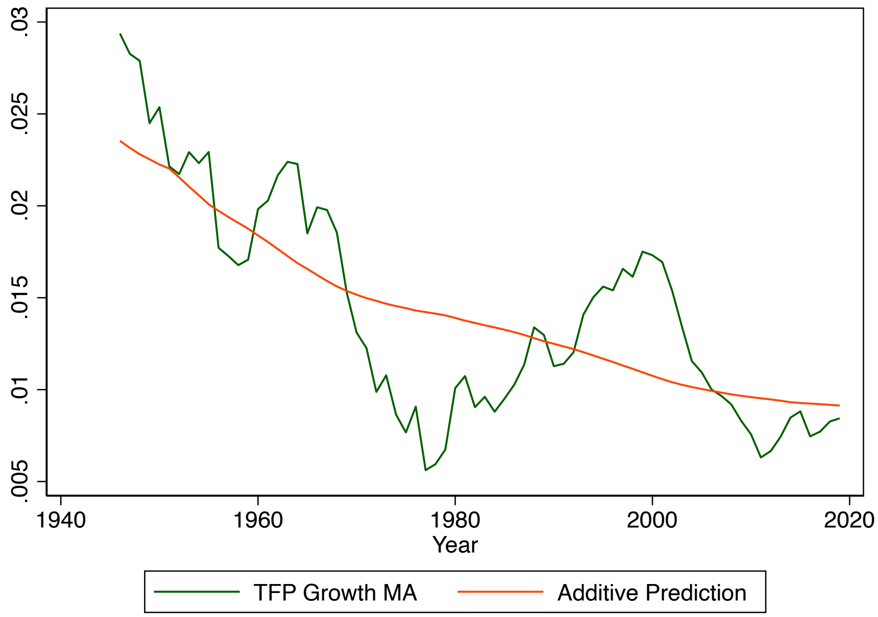
Notes: TFP is normalized to 1 in 1947. Models are estimated over 1947-1983. The forecast 1984-2019 is out-of-sample. Data source: Bergeaud et al. (2016).

theory.

Section 1 focuses on US data and uses a relatively simple test based on the root mean squared errors of forecasts (RMSE) from models $\mathcal{A}$ and $\mathcal{G}$. The test confirms the visual impression of figure 1. Model $\mathcal{G}$ performs poorly even when one allows the trend g to change over time. Model $\mathcal{A}$ forecasts TFP better than model $\mathcal{G}$ at all horizons. The test based on RMSEs rejects model $\mathcal{G}$ with more than 95% confidence even conditionally on a TFP slowdown. The same results hold in the longer sample (1890-2019) of Bergeaud et al. (2016). The long sample also reveals 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 suggests that US TFP growth has been additive with constant expected increments for 90 years.

Section 2 extends the analysis with a Bayesian model. The parameters are estimated by maximum likelihood, filtering the distribution of the unobserved states with a Kalman filter, before approximating the conditional expectations using Monte Carlo simulations. While more complicated than that of Section 1.2, the Bayesian estimation provides optimal forecasts as well as posteriors for model selection. Figure 6 shows the conditional forecasts of the two models and highlights the failures of model $\mathcal{G}$. In the model selection

Figure 2: Predicted TFP Slowdown



Notes: TFP is normalized to 1 in 1947. 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).

exercise the posteriors converge to one for model $\mathcal{A}$ even when one starts with large priors in favor of model $\mathcal{G}$ in 1890.

Section 3 repeats the analysis of Sections 1.2 and 2 in the long panel of 23 countries and 129 years from Bergeaud et al. (2016). The additive model predicts TFP dynamics better than the exponential model for each of the 23 countries. The 10-year forecast errors of the exponential model are 30% to 60% higher than those of the additive model and model selection favors model $\mathcal{A}$ in all cases. I also consider a sample of OECD countries that are not in the BCL sample (e.g., Korea) and I show that their TFP growth is linear. TFP growth paths in Thailand and Taiwan, two prime example of "miracle growth" in Asia, are also linear. The exponential growth model fails because it predicts periods of sustained and convex productivity growth that simply do not exist in the data.

Section 4 provides a broader historical interpretation of the data. A symptom of the failure of the exponential model is that the estimated trend growth rates are unstable. By contrast the additive TFP model displays remarkably few breaks and these have plausible economic interpretations in terms of General Purpose Technologies (Bresnahan and Trajtenberg, 1995). For example, the process of US TFP increments has only one break over the past 130 years, around 1930, following the large-scale implementation of the electricity revolution (Gordon, 2016). Using UK data on GDP per capita back to 1500, I find only two more breaks. The first is around 1650 and the second is around 1830. These breaks are consistent with historical research on the first and second industrial revolutions (Mokyr and Voth, 2010).

Section 5 connects additive growth with general equilibrium theory, in two steps. Additive growth does not substantially change the short-run dynamics of a standard stochastic growth model with exogenous TFP. Wealth effects are somewhat muted in model $\mathcal{A}$ relative to model $\mathcal{G}$, but the policy functions for consumption and investment are quantitatively similar for reasonable values of the elasticity of inter-temporal substitution. Even if agents believe (incorrectly) that the underlying process is exponential, they save (approximately) the same fraction of their income, supply the same quantity of labor, and the economy accumulates the same quantity of capital. The fact that TFP growth is additive therefore has limited implications for monetary policy and business cycles, and for the great ratios (labor share, capital to GDP ratio, etc.) *given* observed TFP. Long-term rates are less volatile in model $\mathcal{A}$ than in model $\mathcal{G}$, however, and important exceptions (debt sustainability and climate change) are discussed in the conclusion.

The second part of Section 5 studies additive TFP in models of endogenous growth (Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992). The important point is that exponential growth derives from the *assumption* that ideas are multiplicative. Models with expanding varieties *assume* that the number of potential new varieties is proportional to the number of existing varieties. Models with vertical differentiation *assume* that the quality ladder itself is exponential. I explain how changing these assumptions leads to additive growth and I discuss the issues that may then arise. I also discuss additive growth in the context of models of search and imitation such as Lucas and Moll (2014), Perla and Tonetti (2014) and Akcigit et al. (2018). An important policy implication of this analysis is that inter-temporal spillovers are smaller under model $\mathcal{A}$ than under model $\mathcal{G}$ since improvements in TFP raise economic efficiency but do make *future* discoveries exponentially easier. The rationale for R&D subsidies is thus weaker under additive growth.

Literature 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. My Bayesian estimation rejects the exponential model using the full sample, but, interestingly, the power of the test is lower if I only use Solow's original data because of the structural break in 1930.

This paper complements the literature on endogenous growth accounting, such as Jones (2002). Compared to Solow (1957), this literature treats TFP growth as an endogenous variable to be explained by inputs such as capital, education, and the labor force employed in research. A key puzzle in the literature is that TFP growth has not increased despite the increase in measured research effort (Jones, 1995). Jones (2009) argues that innovation is getting harder because new generations of innovators face an increasing educational burden. Similarly, 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. A critical issue in this literature is the measurement of inputs into the innovation process. Formal R&D spending captures only a fraction of innovative activities, and R&D data is usually missing before World War 2. Growth accounting also depends on the functional form chosen to map inputs into TFP. By contrast, my approach focuses directly on the TFP process and shows that TFP is additive across many countries and time periods. This new stylized fact speaks to all models of growth, endogenous or semi-endogenous, based on R&D or on learning-by-doing.

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. My results suggest that $\phi = 0$.

Finally this paper relates to the history of long run growth. Transitions between regimes, as in Figure 13, are even more striking when viewed through the lens of additive 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.*" The early break point around 1650 is consistent with recent work by Bouscasse et al. (2021).

1 Evidence from the US

My primary sources for TFP is Bergeaud et al. (2016) (BCL). BCL covers 23 countries from 1890 to 2019 and their data allow the analysis of a long sample period as well as international comparisons in Section 3.

The challenge is to create a test that can distinguish between additive and geometric growth. Consider a simple 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 $A_0^* = 1$ and assume that fundamental TFP is deterministic and that the trend measured at time 0 is b. We do not know whether A_t^* will grow exponentially as $A_{t,\mathcal{G}}^* = e^{bt}$ (model $\mathcal{G}$) , or linearly as $A_{t,\mathcal{A}}^* = 1 + bt$ (model $\mathcal{A}$). How long would we wait before deciding which is the correct model? If we want to the two forecasts to be $n\sigma$ apart we need $e^{bt} - 1 - bt \geq n\sigma$ or $bt \geq x$ 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 $t \geq \frac{\sqrt{2n\sigma}}{b}$. With $n = 2$, $\sigma = 1\%$ and $b = 2\%$ 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 basic insight carries over to the more advanced models used below.

The example above is of course not realistic. The TFP process is probably not simply the sum of a deterministic component and measurement error. The trend b must be estimated and can vary over time. In addition, we want to formalize the test instead of choosing an arbitrary degree n. The following sections answer all these questions. I start with a simple test based on the root mean square error of recursive forecasts. Section 2 presents a full Bayesian estimation and likelihood ratio test.

1.1 Simple Test using Postwar U.S. Data

This section presents a first simple test. I focus first on post-war US data because it is the most reliable and because the US was arguably at the technological frontier during the entire period.

The simple test is based on a forecasting exercise. Given an sequence of observed TFP A_t I construct the log growth rate as $g_t \equiv \log A_t - \log A_{t-1}$, and the linear increment as $b_t \equiv A_t - A_{t-1}$. I then estimate the local drifts

$$\hat{g}_t = \mathbb{E}_t [g_{t+1}] \text{ and } \hat{b}_t = \mathbb{E}_t [b_{t+1}] .$$

The simple test forecasts time varying growth according to a standard smoothing model

$$\hat{b}_t = (1 - \zeta)\,\hat{b}_{t-1} + \zeta b_t, \tag{3}$$

and similarly for $(\hat{g}_t, g_t)$. I use values of 0.05 and 0.1 for the smoothing parameter ζ. At 0.05, the sensitivity of the trend estimate to the most recent observation is the same as that of a 20-year moving average. At 0.1 the sensitivity would be the same as that of a 10-year moving average. These cover the range of plausible values for estimating a trend. Equation (3) implies that g_t and b_t are martingales, so that $\mathbb{E}_t[g_{t+h}] = \mathbb{E}_t[g_{t+1}]$ for all $h > 1$. The T-period ahead forecasts for model $\mathcal{A}$ is therefore

$$\mathbb{E}_t[A_{t+h} \mid \mathcal{A}] = A_t + \hat{b}_t h,$$

and model $\mathcal{G}$

$$\mathbb{E}_t[\log A_{t+h} \mid \mathcal{G}] = \log A_t + \hat{g}_t h.$$

Finally I compute the forecast errors

$$\epsilon^{\mathcal{A}}_{t+h} = \frac{A_{t+h} - \mathbb{E}_t[A_{t+h} \mid \mathcal{A}]}{\bar{A}},$$

where $\bar{A}$ is the sample average of A_t (this normalization eases the comparison across datasets). For model $\mathcal{G}$ the forecast error of log TFP is

$$\epsilon^{\mathcal{G},\log}_{t+h} = \log A_{t+h} - \mathbb{E}_t[\log A_{t+h} \mid \mathcal{A}].$$

Similarly we can compute the log errors of model $\mathcal{A}$, $\epsilon^{\mathcal{A},\log}_{t+h}$, and the level errors of model $\mathcal{G}$, $\epsilon^{\mathcal{G}}_{t+h}$. These forecasts are non-linear and incorporate Jensen inequality terms, but these corrections are small and make no difference in practice.[1]

Table 3 reports the root mean square errors (RMSE) of the forecasts of the two models, for two values of ζ and of h, and for two loss functions, in levels and in logs. In all cases, model $\mathcal{A}$ beats model $\mathcal{G}$. The gap is around 150 basis points at the 10-year horizon and 400 basis points at the 20-year horizon. The relative performance of the two models does not appear to depend on the loss function. To save space I will focus on the simple quadratic loss in levels for the rest of the paper.

[1]In post-war US data the mean growth rate $\bar{g}$ is 1.36% and the volatility of the 10-year moving average is 0.48%. At the 10 year horizon the mean effect is 13.6% while the variance $(1/2\sigma^2_h)$ is only 0.11%.

Table 1: RMSE of TFP Forecasts, U.S. 1947-2019,

Loss Function	$(A_{t+h} - \mathbb{E}_t[A_{t+h}])^2$				$(\log A_{t+h} - \mathbb{E}_t[\log A_{t+h}])^2$			
Smoothing Parameter	$\zeta = 0.05$		$\zeta = 0.1$		$\zeta = 0.05$		$\zeta = 0.1$	
Horizon h (years)	10	20	10	20	10	20	10	20
Model $\mathcal{A}$	.0473	.0694	.0527	.0813	.0448	.0625	.0497	.0734
Model $\mathcal{G}$	.0637	.1123	.0662	.1204	.0586	.1000	.0612	.1072

Notes: US TFP is from the updated work of Bergeaud et al. (2016)

I now wish to test the power of the test with respect to the hypothesis that the data is generated by model $\mathcal{G}$. Is the fact that model model $\mathcal{A}$ beats model $\mathcal{G}$ by 150bps at the 10-year horizon significant or not? Random fluctuations in growth rates can, especially in small samples, generate type-1 errors. How likely are these? Table 2 presents the key statistics for post-war data.

TFP growth is close to iid. The autocorrelation $(\Delta \log A_t, \Delta \log A_{t-1})$ is zero. A simple benchmark, then, is to model growth as

$$g_t = \bar{g} + \epsilon_t,$$

where $\bar{g} = 1.36\%$ and ϵ is white noise with $\sigma_\epsilon = 1.22\%$. To take into account the limited sample size I simulate 5,000 TFP path of 73 years of growth. Figure 3 shows the results of applying the simple RMSE test to the simulated data. The median value of the test is around 65 basis points. As expected it is difficult to distinguish exponential growth in a relatively short sample. In many simulations the difference in RMSE is small and in a quarter of the simulations model $\mathcal{A}$ performs a bit better even though the data is generated by model $\mathcal{G}$. Very few simulations, however, produce differences of more than 100 basis points in favor of model $\mathcal{A}$. The 5th percentile is -1% and the first percentile is -1.8%. If we use the decision rule "*reject model $\mathcal{G}$ if model $\mathcal{A}$ beats it by more than 150bps*", consistent with the evidence in Table 1, the risk of type-1 error is slightly less than 2%.

One might argue, however, that one should perform the test *conditional* on a TFP slowdown. There are 2,495 [2] simulations where average TFP growth from $t = 37$ to $t = 73$ is higher than average TFP growth from $t = 1$ to $t = 36$. Conditional on a TFP slowdown the risk of type 1 error rises to approximately 4%.[3]

[2] The expected value is 2500 (half of 5,000) but there is some sampling noise.

[3] I should emphasize that it is meaningless to test for *unrestricted* TFP slowdowns since this would

Table 2: Monte Carlo Simulations

US Data	Obs.	Mean	Std. Dev.	Min.	Max
Log growth rate: $\Delta \log A_t$	73	.0136	.0122	-.009	.037

Simulated Data	Obs.	Mean	Std. Dev.
Log growth rate: $\Delta \log A_t$	$73 \times 5,000$	.0136	.0122

Test Results	Obs.	Mean	Median	Std. Dev.	5^{th}	1^{th}
$RMSE(\mathcal{A}) - RMSE(\mathcal{G})$	$5,000$	0.6%	0.65%	0.96%	-1.06%	-1.81%

Test Power	Obs.	$\Delta RMSE < 1.5\%$	Type 1 Errors
Full MC Sample	$5,000$	98	1.96%
Sample with TFP Slowdown	$2,495$	98	3.93%

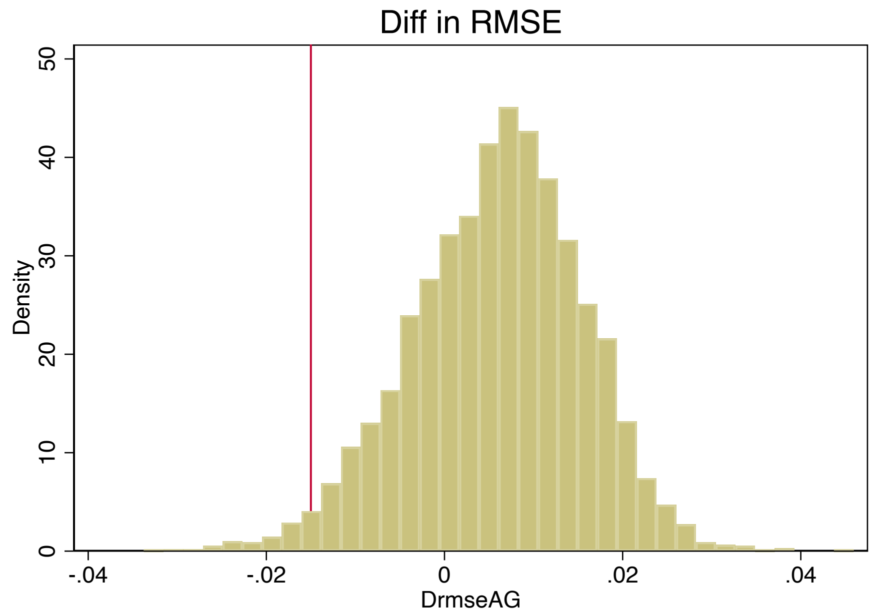
Notes: US TFP 1947-2019 from Bergeaud et al. (2016). Growth rates winsorized at 5%,95%. Simulated data based on 1,000 simulations of 73 years of growth, matching the mean and volatility of US TFP growth. On average model $\mathcal{G}$ beats model $\mathcal{A}$ in the RMSE test by 60 basis points. The risk of type 1-error at 150 basis points (model $\mathcal{A}$ beating model $\mathcal{G}$ by more than 150bps) is 2% in the full sample of 5,000 simulations. There are 2,495 simulations with a TFP slowdown (average TFP growth from t=1 to 36 is higher than from t=37 to t=73). Conditional on a TFP slowdown the risk of type 1 error rises to 3.9%.

Result 1: *Postwar US TFP growth is well described by model $\mathcal{A}$ with constant increments around 250 basis points each year starting from a value of 1 in 1947. Model $\mathcal{A}$ forecasts TFP better than model $\mathcal{G}$ by about 150 basis points at the 10-year horizon. The RMSE test rejects model $\mathcal{G}$ with more than 95% confidence even conditionally on a TFP slowdown.*

As explained in the introduction, this result is robust to the measurement of the capital stock. If TFP is additive then labor productivity – the *product* of TFP and capital intensity – is convex. Panel (b) in Figure 16 shows that the convex-linear forecast of model $\mathcal{A}$ predicts correctly the evolution of labor productivity in the long term. Model $\mathcal{G}$ does not.

actually include model $\mathcal{A}$, where the slowdown is simply proportional to $1/A_t$ as shown in Figure 2. As an alternative I have used estimations with structural breaks instead of the continuous updates in equation (3) and the results are the same. Model $\mathcal{A}$ needs fewer breaks that model $\mathcal{G}$ and its forecasts are more accurate.

Figure 3: RMSE of $\mathcal{A}$ vs $\mathcal{G}$ in $\mathcal{G}$-Simulated Data



Notes: The goal of the figure is to assess the prevalence of type-1 errors in the simple RMSE test, i.e., the risk that the test rejects model $\mathcal{G}$ in favor of model $\mathcal{A}$ even though growth is actually exponential. The data is generated by 10,000 simulations of 73 years of random growth under model $\mathcal{G}$, matching moments of US post war TFP growth. The simple RMSE test is then run on these 10,000 simulated economies. The figure shows the $RMSE(\mathcal{A}) - RMSE(\mathcal{G})$. Since model $\mathcal{G}$ is used to generate the data, we expect a positive difference.

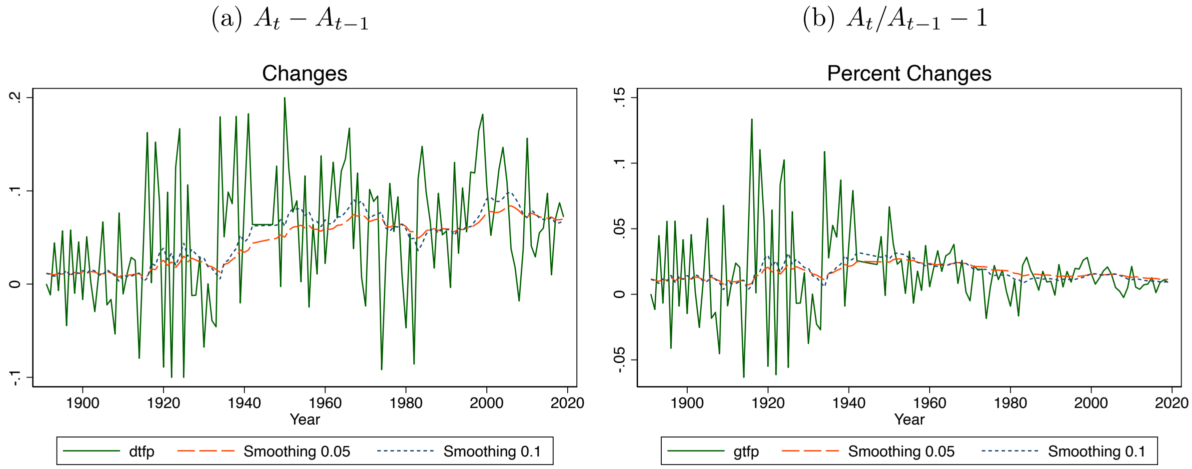
1.2 U.S. 1890-2019

I now extend the sample to pre-WW2 data, focusing again on the simple test based on RMSE. In the next section I present a MLE estimation with Kalman filtering and Bayesian model selection, which is more powerful but significantly more complex.

Figure 4 shows the raw and smoothed series for $\zeta = 0.05$ and $\zeta = 0.1$. The data is from Bergeaud et al. (2016). The model is initiated over the first 10 observations, 1891 to 1900: $\mathbb{E}_{1900}\left[y_{1901}\right] = \frac{y_{1891}+..y_{1900}}{10}$. I drop observations between 1942 and 1946 to avoid extreme outliers during WW2.

A few features stand out in Figure 4. The series in first difference is approximately homoskedastic. The standard deviation of TFP changes is 0.13 before WW2 and 0.11 since 1947, and the difference is not statistically significant. The series is percent changes, on the other hand, displays a secular decline in volatility. The second striking feature is that there is a permanent mean change around 1930. I will return to this point later.

Figure 4: US TFP, 1890-2019



Notes: Models are estimated over 1947-1980. The left panel show the prediction of a linear model. The right panel shows the prediction of a log-linear model. US TFP is from the updated work of Bergeaud et al. (2016).

Table 3: RMSE for US TFP Forecasts, 1890-2019

1890-2019	Levels			
Smoothing Parameter	$\zeta = 0.05$		$\zeta = 0.1$	
Forecast Horizon	10 years	20 years	10 years	20 years
Model $\mathcal{A}$	.086	.145	.090	.147
Model $\mathcal{G}$	.107	.209	.114	.237

Notes: US TFP is from the updated work of Bergeaud et al. (2016)

Forecasts Errors As in the post-war sample, Table 3 shows that model $\mathcal{A}$ outperforms model $\mathcal{G}$ in all cases and the relative performance of model $\mathcal{A}$ *increases* with the forecast horizon.[4] The main reason is that after a sequence of positive growth rates the multiplicative model extrapolates exponential growth for 10 years, which systematically fails to materialize.

Result 2. *For US TFP over 1890-2019, model $\mathcal{G}$'s long-term forecast errors are 25% to 40% higher than those of* model $\mathcal{A}$.

[4]Model $\mathcal{A}$ also beats model $\mathcal{G}$ when RMSE is computed with *relative* errors $\frac{A_t - \mathbb{E}^{b,g}_{t-10}[A_t]}{A_t}$, which corresponds to a loss function based on log-mistakes. This changes the relative importance of mistakes early in the sample versus late in the sample, but model $\mathcal{A}$ continues to outperform.

2 Bayesian Model Selection

This section presents a formal analysis of models $\mathcal{A}$ and $\mathcal{G}$ through the lens of Bayesian model selection. I consider a decision maker (DM) who, faced with some data on TFP, $A^T \equiv \{A_t\}_{t=1}^T$, must choose between model $\mathcal{A}$ and model $\mathcal{G}$.[5] Define the decision maker's priors $p(\mathcal{A})$ and $p(\mathcal{G}) = 1 - p(\mathcal{A})$, and the likelihood of the data given the model $\mathcal{M} = \mathcal{A}, \mathcal{G}$ as $f\left(A^T \mid \mathcal{M}\right)$. The DM can then use Bayes' rule to calculate the posterior probability of the model conditional on the data. In the spirit of the previous section, I take this calculation one-step further and calculate posterior probabilities at the forecasting horizon h:

$$\pi^h\left(\mathcal{M} \mid A^T\right) = \frac{f^h\left(A^T \mid \mathcal{M}\right) p(\mathcal{M})}{f^h\left(A^T \mid \mathcal{A}\right) p(\mathcal{A}) + f^h\left(A^T \mid \mathcal{G}\right) p(\mathcal{G})} \tag{4}$$

I will consider various values for h.

2.1 Hidden Markov Models

To take into account changing growth rates, I cast models $\mathcal{A}$ and $\mathcal{G}$ within a hidden Markov Chain framework with time-varying parameters. I specify model $\mathcal{A}$ as

$$b_t = b_{t-1} + \sigma_u u_t \tag{5}$$

$$A_t = A_{t-1} + b_t + \sigma_a \epsilon_t^a \tag{6}$$

where u_t and ϵ_t^a are uncorrelated, *iid* standard Gaussian innovations. In the language of Kalman filtering, equation (5) is the *state* equation. The state is the (unobserved, latent) trend b_t subject to (unobserved) shocks u_t. Equation (6) is the observation equation. I observe TFP A_t at time t. The change in TFP from $t-1$ to t reflects the sum of the underlying trend and the temporary shock ϵ_t^a, which includes measurement errors. Similarly, the state and observation equations of model $\mathcal{G}$ are

$$g_t = g_{t-1} + \sigma_\nu \nu_t \tag{7}$$

$$A_t = A_{t-1}\left(1 + g_t\right) + \sigma_g \epsilon_t^g \tag{8}$$

[5]This is a binary choice. Andrew et al. (2021) explains how a fully Bayesian approach can encounter difficulties for problems with a continuum of choices, (e.g. if the DM could choose a linear combination of models A and G).

where ν_t and ϵ_t^g are uncorrelated, *iid* standard Gaussian innovations.

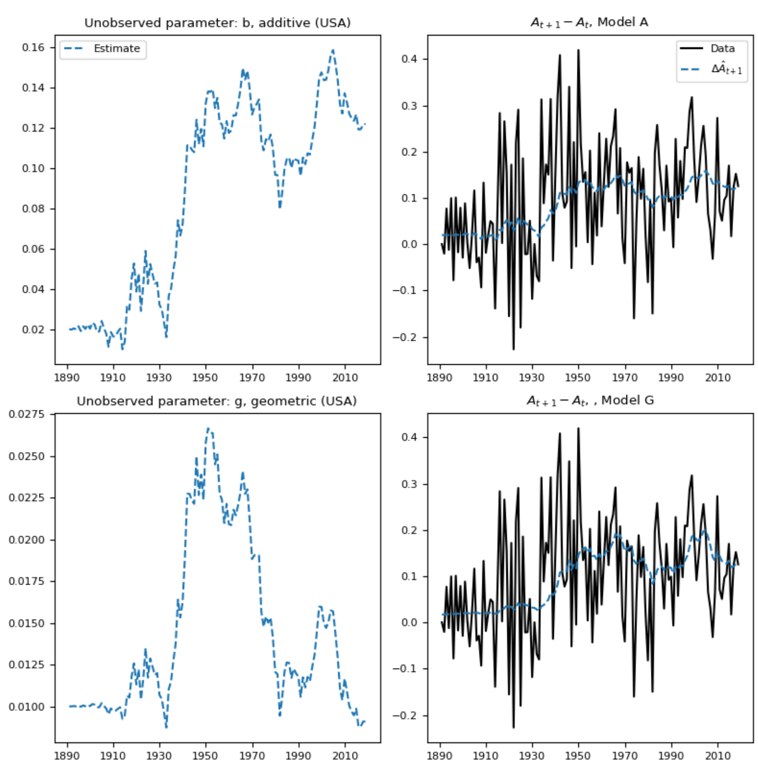
2.2 Estimation and Inference

I estimate the parameters $\theta_A = \{\sigma_u, \sigma_a\}$ and $\theta_B = \{\sigma_\nu, \sigma_g\}$ via maximum likelihood, filtering the distribution of the unobserved states via the Kalman filter. The innovations representation of the Kalman filter (Ljungqvist and Sargent, 2018) and the distributions of the latent states in each model, $\{b_t\}$ and $\{g_t\}$, are estimated conditional on the parameters.

I let the Kalman filter estimate the data from 1942 to 1946 to avoid outliers (Durbin and Koopman, 2008). Equations (5) and (7) assume that the trends are random walk. I confirm that the residuals are indeed close to *iid*, with small autocorrelations of $-.09$ for model $\mathcal{A}$ and $-.002$ for model $\mathcal{G}$.

Figure 5 shows the filtered estimates $\hat{b}_t = \mathbb{E}\left[b_{t+1} \mid A^t\right]$ and $\hat{g}_t = \mathbb{E}\left[g_{t+1} \mid A^t\right]$, where $A^t = (A_t, \ldots A_0)$ denotes the history up to time t. In the top panel, I show the mean estimate of b_t (top left), and how it relates to the data (top right). The estimates confirm the large increase in $\hat{b}_t$ around the 1930s from Figure 4.

Figure 5: Estimated unobserved coefficients for US TFP



The bottom panel of Figure 5 shows $\hat{g}_t$ and how it relates to the data (bottom right). We see again the large increase in the growth rate around 1930 but unlike model $\mathcal{A}$ the increase is transitory.

2.3 Conditional Forecasts

With parameter estimates in hand I can calculate the conditional forecasts of $A_{t+h} \mid A^t$ for any horizon h. Figure 6 shows the conditional forecasts of Models $\mathcal{A}$ and $\mathcal{G}$ for the US economy between 1890 and 2019. For model $\mathcal{A}$ the conditional mean is simply $\mathbb{E}\left[A_{t+h} \mid A^t\right] = A_t + h\hat{b}_t$. For model $\mathcal{G}$, however, I must approximate the conditional expectations using Monte Carlo simulations. The technical details are in Appendix B.

The top panel of Figure 6 plots the term structure TFP forecasts implied by model $\mathcal{A}$, conditional on the estimated parameter vector $\hat{\theta}_{\mathcal{A}} = \{\sigma_u, \sigma_a\}$ and the filtered state

$$\mathbb{E}\left[A_{t+h} \mid A_t, \hat{\theta}_{\mathcal{A}}, \hat{b}_t; \mathcal{A},\right]$$

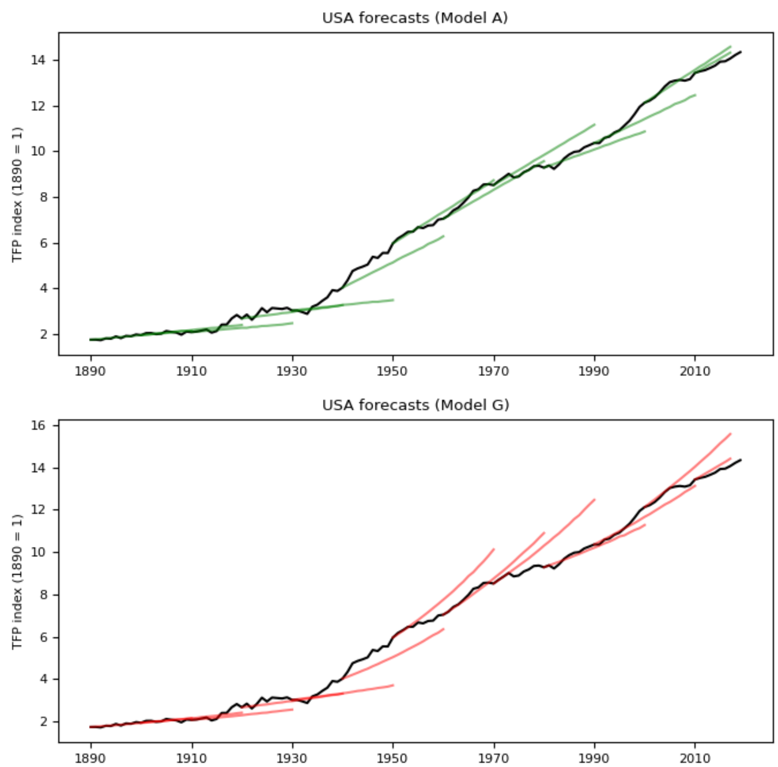
We observe large forecast errors around the 1930s, inline with the large change in $\hat{b}_t$ discussed earlier. Once the filtered estimates of b_t adjust the forecasts errors do not appear substantial, even at longer horizons.

The bottom panel of Figure 6 shows the conditional expectations from model $\mathcal{G}$, $\mathbb{E}\left[A_{t+h} \mid A_t, \hat{\theta}_{\mathcal{G}}, \hat{g}_t; \mathcal{G},\right]$. These conditional expectations are computed by MC simulations because of the non-linearities in model $\mathcal{G}$. We again observe large forecast errors around the 1930s but even once the filter adjusts the exponential nature of model $\mathcal{G}$ leads to large forecast errors at long horizons. The next section formalizes this intuition.

2.4 Model Comparison

The main advantage of Bayesian estimation lies in its ability to formally compare competing models. To do so I calculate the posterior probabilities in equation 4 starting with the conditional likelihoods at different forecast horizons, h. While full Bayesian estimation would require $f^h\left(A^T \mid \mathcal{M}\right) = \int f^h\left(A^T \mid \theta_{\mathcal{M}}\right) d\pi\left(\theta_{\mathcal{M}}\right)$, the well-documented numerical instabilities involved in calculating marginal densities motivate the use of the maximum likelihood estimate $f^h\left(A^T \mid \mathcal{M}\right) = f^h\left(A^T \mid \theta_{\mathcal{M}}^{MLE}\right)$. The simplifying assumption that the estimates $\hat{b}_t$ and $\hat{g}_t$ are known with certainty implies that the conditional forecasts at any horizon are Gaussian, substantially reducing the complexity of the likelihood calculation. It also implies that the distribution of the vector of forecasts is fully

Figure 6: Conditional forecasts from time-varying parameter model, USA



characterized by its mean and covariance matrix.

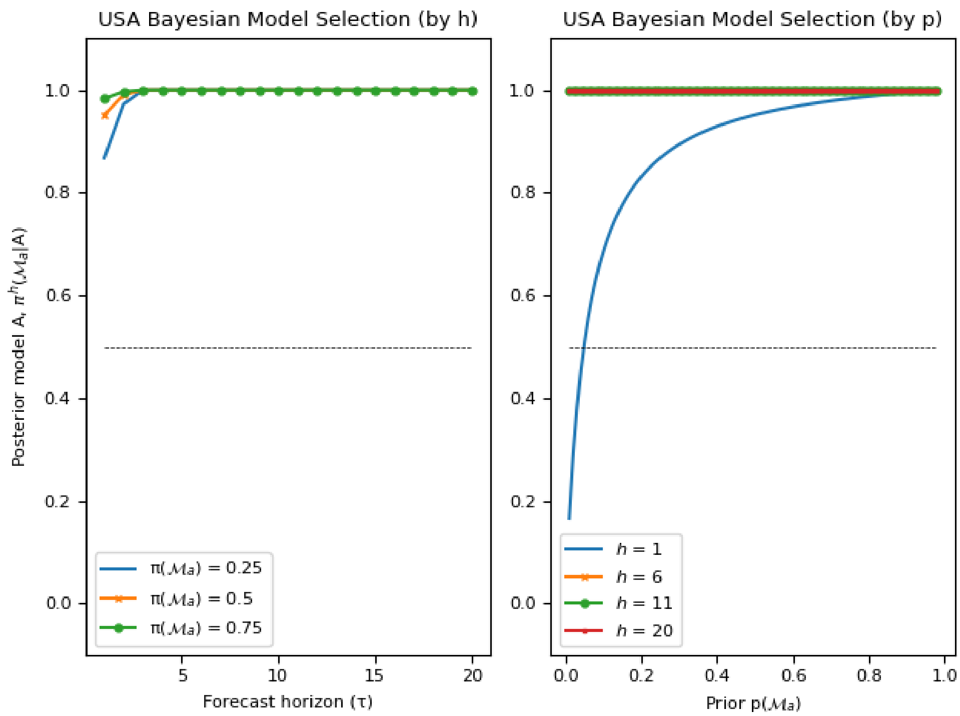
The h-steps ahead likelihood are evaluated recursively. Since each density is Gaussian it is enough to evaluate the mean and variance of the conditional forecasts for both models. The conditional variance of model $\mathcal{A}$ is a (relatively) simple recursive equation involving the Kalman gains and the conditional one-step variances. For model $\mathcal{G}$, I must again use Monte Carlo simulations described in Appendix B. Moreover, since the forecasts are serially correlated, I must calculate the likelihood of the vector of forecasts in its entirety.[6]

The left panel of Figure 7 shows the posterior probability of Model $\mathcal{A}$, $\pi^h\left(\mathcal{A} \mid A^T\right)$ as a function of the forecast horizons τ. The different colored lines correspond to different priors of the decision maker. The right panel of Figure 7 slices the same data in a

[6]In practice we use MC simulations for model $\mathcal{A}$ as well as for model $\mathcal{G}$ to ensure that the two models are treated in exactly the same way. We then use the recursive formula available for model $\mathcal{A}$ to verify the accuracy of the MC simulations.

different way, showing the posterior probability as a function of prior probabilities for different forecast horizons h. The posterior is one irrespective of the prior for all horizon except $h = 1$.

Figure 7: Posterior Selection Probability for Model $\mathcal{A}$, USA



Result 3. *The Bayesian model select* model $\mathcal{A}$ *with posterior probability of one in the long US Sample.*

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 6. 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, as I show in Section 4, a formal structural test finds a 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. The Bayesian approach therefore also explains why past research might have concluded that model $\mathcal{G}$ was appropriate.

3 International Evidence

Bergeaud et al. (2016) provide data for 23 countries.[7] I now apply the models of Section 1.2 to each country. This exercise is useful for two reasons. The first reason is that the sample is much larger. The second reason is that I can investigate TFP growth in countries that are not at the frontier of technology. I find that model $\mathcal{A}$ beats model $\mathcal{G}$ in *all* countries, and often by a wider margin that in the US. Interestingly, catch-up growth is also (conditionally) linear.

3.1 Simple Model

The trend growths are estimated with the recursive learning model (3) with parameter $\zeta = 0.05$ and $\zeta = 0.1$. As before, all the forecasts are out-of-sample. For each country $i = 1 : 23$ and each year t I compute the forecast errors at the 10 year horizon as

$$\epsilon_{i,t}^{\mathcal{M}} = \frac{A_{i,t} - \mathbb{E}_{t-10}^{\mathcal{M}}\left[A_{i,t}\right]}{\bar{A}_i},$$

where $\bar{A}_i$ is the country sample average and the expectation are taken under models $\mathcal{A}$ and $\mathcal{G}$. Finally, I compute the root mean square error for each country as

$$\texttt{RMSE}_i\left(\mathcal{M}\right) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\epsilon_{i,t}^{\mathcal{M}}\right)^2}.$$

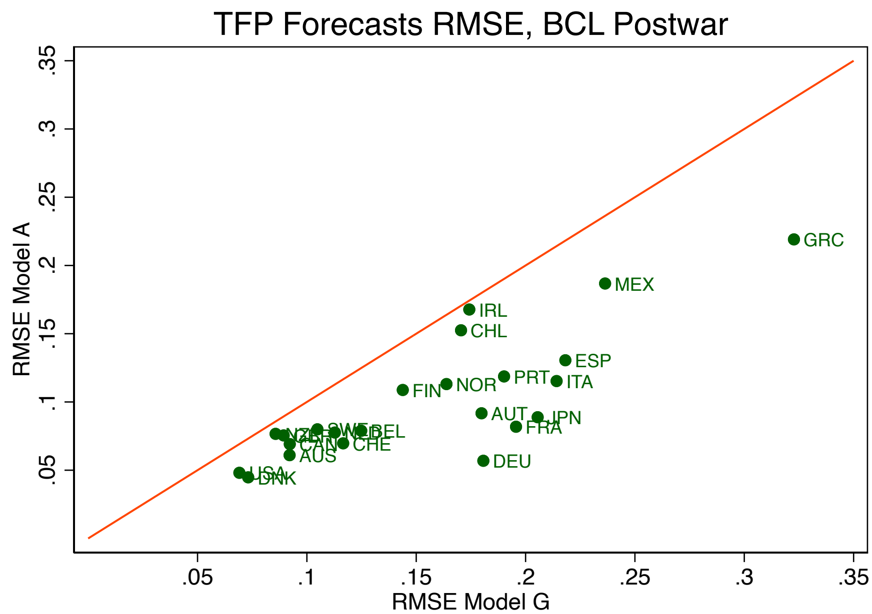
I consider first the 1950-2019 sample to avoid the disruption of World War II.[8] Figure 8(a) shows the RMSE of TFP forecasts in the two datasets. Model $\mathcal{A}$ performs better than model $\mathcal{G}$ in all cases. In most cases the relative performance of model $\mathcal{A}$ is stronger than for the US.

[7] 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 United States. The sample covers 1890–2019. The main variables are GDP, labor, and capital. Labor is constructed from data on total employment and working time. Capital is constructed by the perpetual inventory method applied to equipment and buildings.

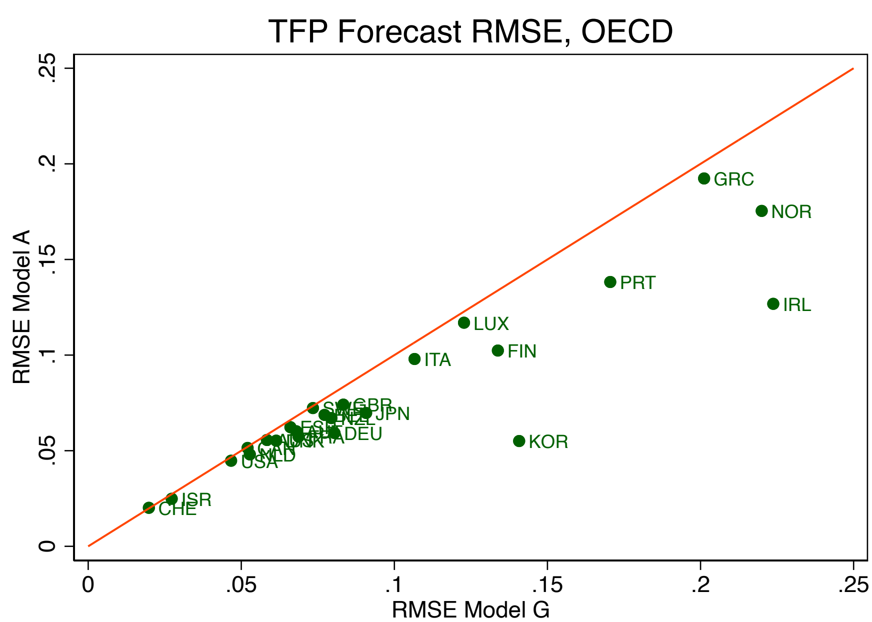
[8] The existence of disasters is a good reason to use TFP as opposed to GDP per capita to study long term growth. In the US, for instance, modern TFP growth starts in the middle of the great depression: a macro disaster coincides with a technological miracle. TFP captures it correctly because it accounts for changes in hours worked. The case of Europe is more complicated because the shock (WW2) is larger and destroys the capital stock, where some of the technology is embedded. It therefore seems safer to first estimate the model in the post-1950 sample. In any case, Table 4 shows that the results are similar in the full sample 1890-2019.

Figure 8: TFP Forecast Errors, Post War

(a) BCL 1950-2019

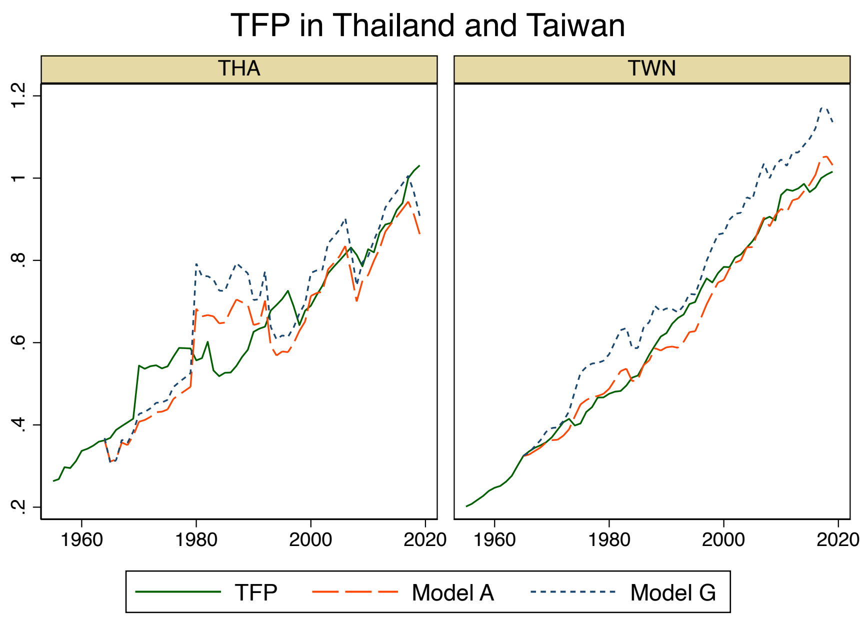


(b) OECD, post-1985



Notes: Model $\mathcal{G}$ on the horizontal axis, model $\mathcal{A}$ on the vertical axis. Out-of-sample, 10-year forecasts with smoothing parameter 0.05. Data from Bergeaud et al. (2016). Sample 1950-2019.

Figure 9: TFP, Fast Growing Asian Countries



Notes: Data from Penn tables Asia. The solid green line is the level of TFP, A_t. The dashed and dotted lines show the forecast made 10 years before by the two models, $\mathbb{E}^{\mathcal{M}}_{t-10}[A_t]$.

I also use the OECD MFP database as a robustness check in Figure 8(b). The data covers 24 countries and starts in 1985 for most, and later for some. Because the time series are much shorter it is more difficult to tell the models apart and some countries are bunched close to the 45 degree line. Nevertheless, model $\mathcal{G}$ never performs better than model $\mathcal{A}$, and often performs worse. Perhaps the most interesting case is that of Korea, which is not in the BCL sample and has experienced strong growth over the past 30 years. It turns out that Korean TFP growth is very linear.

The BCL and OECD data do not include some important Asian countries with strong growth performance. Figure 9 shows TFP for Thailand and Taiwan along with the 10-year forecasts of the two models. Taiwan's TFP growth is remarkable. The TFP index, normalized to 1 in 2017, was only 0.2 in 1955. Such a fast growth makes it easy to tell apart the two models. Model $\mathcal{A}$ fits very well. Model $\mathcal{G}$ vastly over-predicts TFP, irrespective the smoothing parameter.

Table 4 summarizes the average performance of models $\mathcal{A}$ and $\mathcal{G}$. The differences are even larger than in Table 3. Model $\mathcal{A}$ over-performs model $\mathcal{G}$ by 30% to 60%. The table shows that this result also holds for the long sample.

Linear growth inside the frontier could suggest that the limit on productivity growth

Table 4: Average RMSE for 23 Countries, BCL Sample

Sample	1890-2019		1950-2019	
Parameter	$\zeta = 0.05$	$\zeta = 0.1$	$\zeta = 0.05$	$\zeta = 0.1$
Model $\mathcal{A}$	.130	.128	.102	.103
Model $\mathcal{G}$	.171	.168	.162	.145
Number of Countries	23	23	23	23

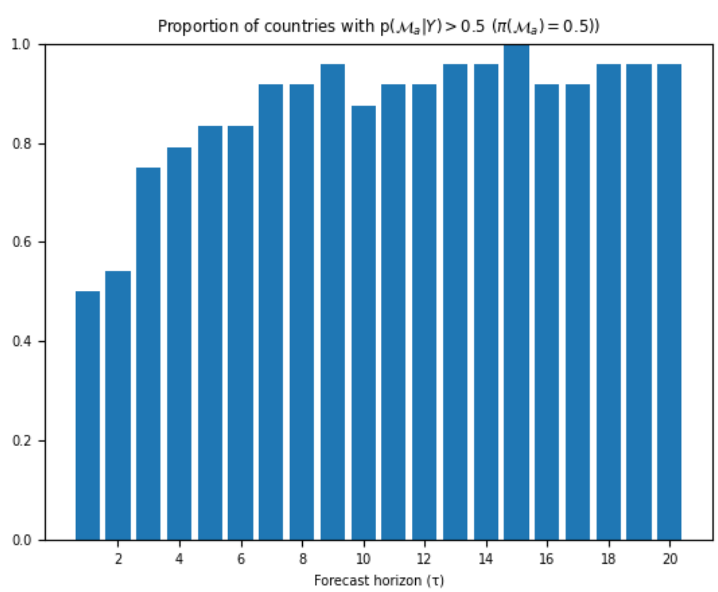
Notes: Data from Bergeaud et al. (2016).

is not the flow of new ideas, but rather their implementation, via human capital investment and learning-by-doing (Comin and Hobijn, 2010).

3.2 Bayesian Selection

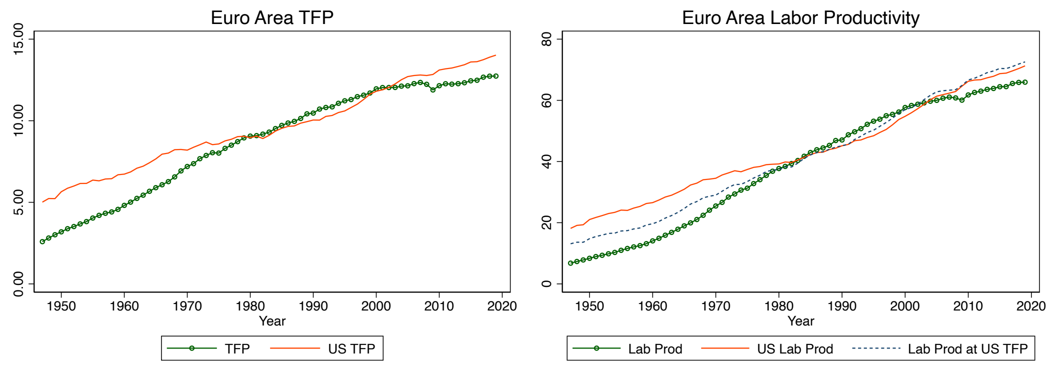
I run the same model selection methodology used for the US for all the countries in the BCL sample. Figure 10 shows the proportion of countries with posterior probability of Model $\mathcal{A}$ greater than 0.5. The Bayesian model selects model $\mathcal{A}$ for essentially all countries in the sample. The one exception is Ireland where both models fit poorly and there is no clear winner.

Figure 10: Proportion of countries with $p(\mathcal{A}|Data) > 0.5$



Result 4. *TFP growth is better described by* model $\mathcal{A}$ *than by* model $\mathcal{G}$ *for both*

Figure 11: TFP and Labor Productivity in the US and the Euro Area



Sources: BCL. In the right panel, labor productivity is GDP per hour. Using equation (24) EA labor productivity at US TFP is defined as $\tilde{\lambda}_{EA,t} = A_{US,t} k_{EA,t}^{\alpha}$.

developed and developing countries.

3.3 TFP Slowdown Revisited

Section 1.1 has shown that there is no TFP slowdown in the US if one defines the benchmark model as linear. There is, however, a TFP slowdown in the euro area (EA, defined as of current membership) and in Japan. Figure 11 compares the evolution of TFP (left panel) and labor productivity (right panel) in the US and the EA during the post-war era. The EA catches up with the US from 1947 to 1980. Between 1980 and 1990 TFP growth is somewhat faster in the EA than in the US. From the mid 1990s onward, however, EA TFP starts to fall behind US TFP. The right panel shows the same pattern for labor productivity, and the dashed line shows that the entire slowdown in output per hour in the EA comes from TFP, not from capital accumulation.

Table 5 shows the evolution of TFP increments before and after 1991, a year chosen because it corresponds to the peak of the EA relative TFP performance. The increments are scaled by US TFP in 1947 to make them comparable across regions. As I have already shown, the US grows with a roughly constant increment. Until 1990 the EA and Japan grow at a faster pace. After 1990, however, their TFP increments decline dramatically and fall below that of the US. The key point is that this slowdown goes beyond what one might expect at the end of the catch-up period.

Table 5 suggests that model $\mathcal{A}$ provides a more useful benchmark to study growth than model $\mathcal{G}$ does. Model $\mathcal{G}$ is a poor diagnostic tool because no country is able to live

Table 5: TFP Increments

$\Delta\left[TFP\right]/TFP_{US,1947}$	1947-1990	1991-2019
USA	.023	.027
Euro Area	.037	.016
Japan	.029	.012
Denmark	.026	.026
Sweden	.022	.028

Notes: TFP Increments measured in units of US TFP in 1947: $\Delta\left[TFP\right]/TFP_{US,1947}$. Data from Bergeaud et al. (2016).

up to its extreme predictions.[9] If one takes the benchmark to be one of exponentially growing increments, then one must conclude that all countries have failed to live up to our expectations. By contrast the linear TFP benchmark highlights that Denmark and Sweden have TFP performances comparable to that of the US while Japan and the Euro area do not. The exponential model paints a particularly misleading picture of growth in countries with high TFP. The danish TFP increment, at 0.026 is much higher than that of the Euro Area at 0.016. The difference in growth *rates* is less impressive, however, because TFP in Denmark is higher than in the EA.

4 Long Term Historical Evidence

In this section I provide historical evidence of changes in TFP growth in the very long run. In doing so I depart from the formal statistical approach of the previous sections as the historical discussion is more easily framed in terms of regimes separated by breaks.

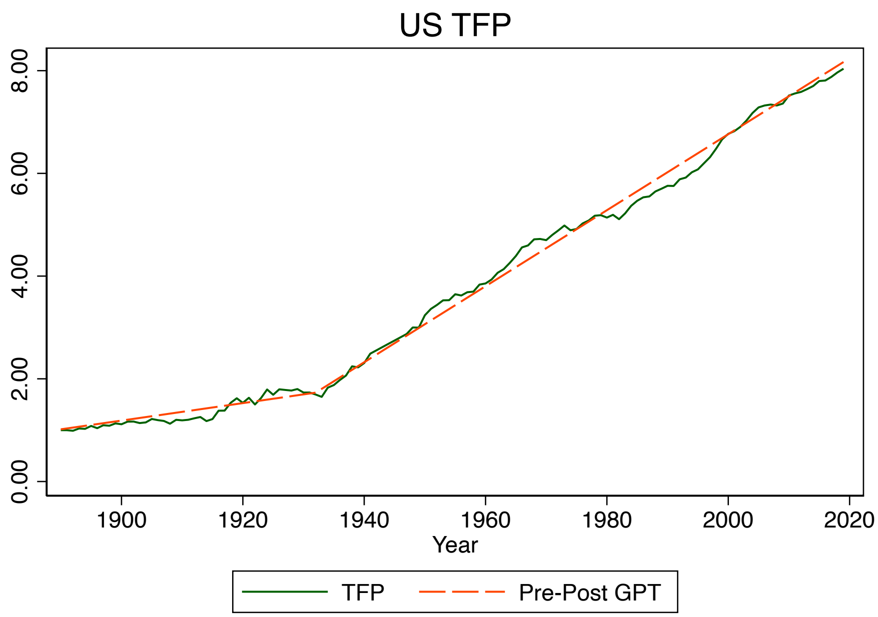
4.1 The 1930 Structural Break

Figure 4 shows that model $\mathcal{A}$, unlike model $\mathcal{G}$, appears to have only one break over the period 1890-2019 in the US. I can formally test this idea following Bai and Perron (2003). The unconstrained test finds one break in the $\Delta\left[TFP\right]$ series: the point estimate is 1933. I test {H0: no breaks} versus {H1: break in 1933}. The W statistic is 21.72 and the p-value is 0.0.

The timing of the break is consistent with Field (2003)'s argument that "*the years 1929–1941 were, in the aggregate, the most technologically progressive of any comparable*

[9]Unless the predicted growth rate is continuously revised downward so as to emulate a linear growth model, but that merely proves the point that the exponential benchmark is useless.

Figure 12: Linear US TFP with One Break



Notes: US TFP is from the updated work of Bergeaud et al. (2016), normalized to 1 in 1890.

period in U.S. economic history." This period corresponds to the large scale implementation of the discoveries of the second industrial revolution: electric light, electric power, and the internal combustion engine, as discussed in Jovanovic and Rousseau (2005). Gordon (2016) points out that it is somewhat surprising that *"much of the progress occurred between 1928 and 1950,"* several decades after the discoveries were made. Following David (1990), he explains the paradox by showing that the 1930s were a period of follow-on inventions, such as the perfection of the piston-powered aircraft and the improving quality of machinery made possible by a large increase in available horsepowers and kilowatt-hours of electricity.

4.2 World TFP: 1550-2020

Just as the US provides a good proxy for the world technological frontier in the 20th and 21st centuries, the UK arguably provides a good proxy in previous centuries. The Maddison series for UK GDP per capita has one observation in the year 1000 and then offers annual values from 1252 onward. Growth appears virtually null until the 1600's (Bolt and van Zanden, 2020). In the neoclassical growth model, labor productivity is proportional to $A_t^{\frac{1}{1-\alpha}}$. If hours worked per capita are stationary and if the capital share is constant then I can use series on GDP per capita to construct proxies for TFP. I make these heroic assumptions and use as my proxy for TFP $(y_t)^{1-\alpha}$ where y_t is GDP per capita and $\alpha = 1/3$.

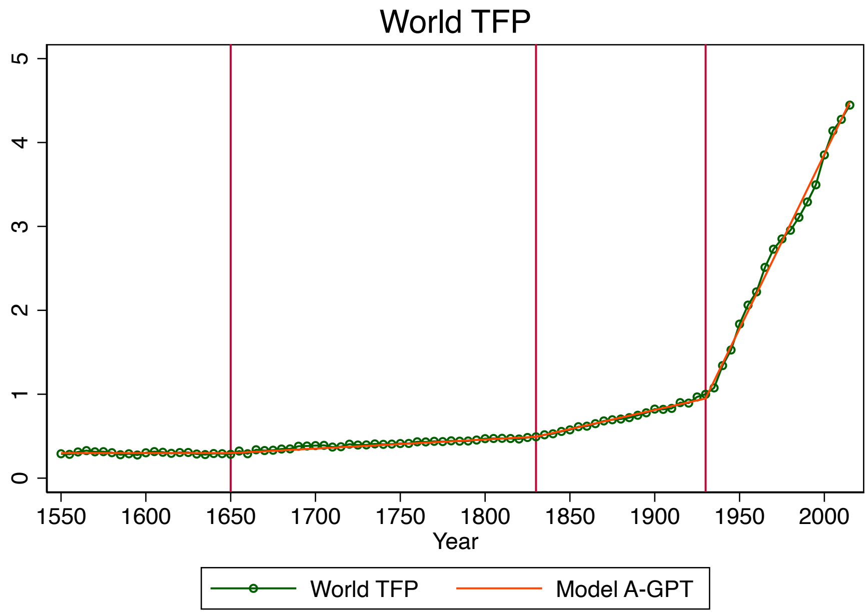
I will use this measure of UK pseudo-TFP for the first part of the sample and then the data from BCL, which start in 1890. An important choice is when to switch from the UK to the US as proxy for the TFP frontier. In the BCL data, the US overtakes Britain in 1910 for GDP per capita but only in the late 1930s for TFP (the US has a higher capital intensity than the UK during that period, which explains the difference). I use 1930 as a switching point. My proxy for World TFP is thus based on the UK before 1930 and on the US after 1930.

Panel (a) of Figure 13 shows the long series for the frontier of World TFP – normalized to 1 in 1930 – together with historical breaks. Each circle represents a five year average. The breaks are in 1650, 1830 and 1930. Panel (b) zooms in on the two main sub-period, 1550-1915 and 1820-2020. Growth is zero until 1650 and the level of TFP is 0.3. Starting in 1650 it increases by 10.5 basis points each year until 1830 where it reaches approximately 0.5. In 1830 the increment increases to 46 basis points and grows linearly until World War 1. The period 1915-1930 is somewhat noisy but TFP remains close to its linear trend of 45 basis points until 1930. After 1930 I observe an enormous increase in TFP growth, from 46 basis points to 418 basis points per year. As a result, TFP today is almost five times higher than it was in 1930.

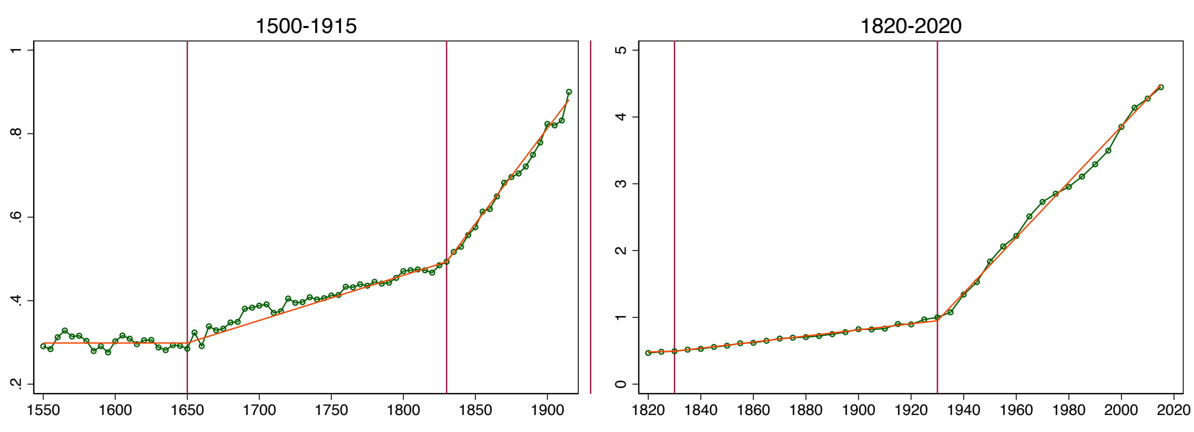
I have discussed the break in 1930 in the previous section. The break in 1830 also aligns well with the standard historical account of the second industrial revolution. The break in 1650 seems to happen before the first industrial revolution, however. There are several explanations for the fact that growth in the UK started earlier than the 18th century. The first key point to keep in mind is that I do not have a measure of hours worked. The pseudo-TFP series are based on income per-capita. Voth (2001) has shown that a rising labor input was an important contributor to growth after 1770. It is plausible that changes in hours per capita also contributed to growth during the previous century. Mokyr and Voth (2010) point out that "*the rise of cottage industries in the countryside after 1650, the famed "proto-industrialization" phenomenon, would do exactly that. There is also reasonable evidence to believe that labor participation rates were rising in the century before the Industrial Revolution.*" Moreover, England, unlike France, had no food crises between 1650 and 1725. Finally, the increase in GPP per capita in the 1600's is consistent with recent work by Bouscasse et al. (2021).

Figure 13: World Frontier TFP

(a) 1600-2020



(b) Sub-periods



Notes: data from Maddison Project. Each circle is an average of 5 years. Pseudo-TFP is $(y_t)^{1-\alpha}$ where y_t is GDP per capita and $\alpha = 1/3$. World frontier is UK until 1930 and US after 1930. TFP normalized to 1 in 1930.

5 Theory

This section discusses the theoretical implications of additive growth, first with exogenous TFP, and then with endogenous TFP.

5.1 Additive Neoclassical Growth

Consider the textbook neoclassical growth model with exogenous TFP. Time is continuous and the labor supply is inelastic. Aggregate value added (GDP, Y_t) is given by

$$Y_t = F\left(K_t, A_t L_t\right) \tag{9}$$

where A_t is labor augmenting (Hicks-neutral) productivity,[10] L_t is the flow of labor services, and K_t is the capital stock which accumulates as

$$\dot{K}_t = I_t - \delta K_t. \tag{10}$$

Labor grows at the constant population growth rate g_L: $\frac{dL_t}{dt} = g_L L_t$. Households have standard CRRA preferences with relative risk aversion γ and rate of time preference ρ. Normalizing the macro variables per efficiency unit of labor as $\hat{k}_t \equiv \frac{K_t}{A_t L_t}$ and $\hat{c}_t \equiv \frac{C_t}{L_t A_t}$, the resource constraint is

$$\dot{\hat{k}}_t = f\left(\hat{k}_t\right) - \hat{c}_t - \left(\delta + g_L + \frac{\dot{A}_t}{A_t}\right)\hat{k}_t, \tag{11}$$

and the Euler equation is

$$\gamma \frac{\dot{\hat{c}}_t}{\hat{c}_t} = f'\left(\hat{k}_t\right) - \delta - \rho - \gamma \frac{\dot{A}_t}{A_t}. \tag{12}$$

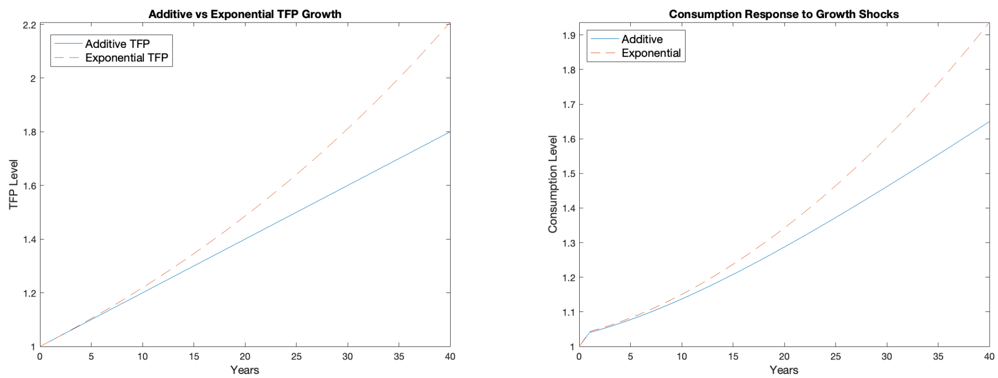
Defining long term TFP growth as $g_{A\infty} \equiv \lim_{t\to\infty} \frac{\dot{A}_t}{A_t}$, the long-term balanced growth path is given by

$$f'\left(\hat{k}_\infty\right) = \delta + \rho + \gamma g_{A\infty}$$

[10] Note that the empirical analysis in Fernald (2012) and Bergeaud et al. (2016) uses standard growth accounting notations where $Y_t = A_t^S K_t^\alpha L_t^{1-\alpha}$ where A_t^S is the Solow residual. In the Cobb-Douglass case there is of course the equivalence $A_t^S = A_t^{1-\alpha}$. Whether one assumes that A_t^S or A_t is linear is of course important empirically but it makes no difference to the theory. I therefore work with A linear because it simplifies the notations.

Figure 14: Response to One Time Increase in Trend Growth

(a) Expected TFP (b) Expected Consumption



Notes: Economy in steady state with zero growth at time 0. The shock is 2% trend growth starting at time 1. The key parameters are CRRA=2, Frisch elasticity=1/2 and capital adjustment costs of 5 (annual frequency) as in $I/K - \delta = (Q-1)/5$. See Appendix for details.

and

$$\hat{c}_\infty = f\left(\hat{k}_\infty\right) - (\delta + g_L + g_{A\infty})\,\hat{k}_\infty$$

In the long run all per capital variables grow with A_t. For instance, long run per capita consumption is $c_t = C_t/L_t = \hat{c}_\infty A_t$. Under exponential growth we have $g_{A\infty} > 0$. There is nothing particularly surprising about the behavior of an economy under permanent additive growth: Growth continues forever, consumption is unbounded but we simply have $g_{A\infty} = 0$ and the risk free rate falls over time towards ρ. Moreover, any growth process less extreme than the exponential one would also have $g_{A\infty} = 0$. Groth et al. (2010) discuss various specifications where this happens and Kruse-Andersen (2022) provides estimates in favor of semi-endogenous growth models.

Since TFP is exogenous in this model we cannot question its path, but we can ask if *given an observed path for TFP*, macroeconomic variables would be different under models $\mathcal{A}$ and $\mathcal{G}$. In other words, we can ask whether agents react differently to the same observed changes in TFP depending on the model they use to interpret the data. To understand whether *policy functions* depend on model specification I simulate the responses of the economy to additive versus exponential trend growth shocks. For the simulations I use a standard discrete time model with elastic labor supply and capital adjustment costs (details are in the Appendix). The important parameters are the CRRA ($\gamma = 2$) and the Frisch elasticity (0.5).

Figure 14 shows the response of the economy to a one-time increase in trend growth.

Table 6: Initial Response to TFP Trend Shocks

CRRA γ		0.5	1	2	3	4	5
Consumption $\frac{C_1}{C_0}-1$ (%)	Model $\mathcal{A}$	0.16	1.86	4.31	6.01	7.29	8.30
	Model $\mathcal{G}$	0.04	1.92	4.66	6.60	8.07	9.24
Short Rate $r_1(1)-r_0(1)$(%)	Model $\mathcal{A}$	0.72	1.16	1.71	2.05	2.32	2.55
	Model $\mathcal{G}$	0.71	1.15	1.68	2.02	2.30	2.54
Long Rate $r_1(\tau)-r_0(\tau)$(%)	Model $\mathcal{A}$	0.79	1.47	2.68	3.77	4.79	5.79
	Model $\mathcal{G}$	0.93	1.75	3.24	4.60	5.91	7.19

Notes: Economy in steady state with zero growth at time 0. The shock is 2% trend growth starting at time 1. The key parameters are CRRA=2, Frisch elasticity=1/2 and capital adjustment costs of 5. The long rate is computed for $\tau = 20$ years. See Appendix for details.

The economy is in steady state with no growth until time 0, with TFP normalized to $A_0 = 1$. Agents wake up at time 1 and observe an increase in TFP from 1 to 1.02. Agents believe the trend increase is permanent. In model $\mathcal{G}$ the agents expect $g = 2\%$ for ever. In model $\mathcal{A}$ they anticipate $b = 0.02$ for ever. Panel (a) shows the expected paths of TFP, one additive, and one exponential. Panel (b) shows the expected path of consumption and I am interested in the *initial* consumption response C_1/C_0 under models $\mathcal{A}$ and $\mathcal{G}$. Note that current TFP is $A_1 = 1.02$ in both cases and the only difference between models $\mathcal{A}$ and $\mathcal{G}$ comes from the expected path of TFP from time $t = 2$ onwards. Consumption jumps on impact because of a wealth effect. The wealth effect could be stronger in model $\mathcal{G}$ than in model $\mathcal{A}$ because agents anticipate higher consumption in the future. The key point, however, is that the increase in consumption in year 1 is quantitatively similar in the two economies. In other words, conditional on the same capital stock and the same observed TFP, agents choose roughly the same level of consumption whether they believe growth to be linear or exponential. Formally, the policy function $C_1 = \mathscr{C}(A_1, K_0; \mathcal{M})$ does not depend much on the model $\mathcal{M}$ in the agents' information set for given state (A_1, K_0). This is a quantitative result that depends on the assumed EIS and figure 14 uses $\gamma^{-1} = 0.5$ as a benchmark. Table 6 shows the robustness of this result.

Table 6 shows the initial response of consumption and interest rates to trend growth shocks for different values of γ. The case $\gamma = 2$ is the one depicted on Figure 14 where consumption increases by 4.3% when agents believe model $\mathcal{A}$ and 4.6% when they believe model $\mathcal{G}$. When $\gamma = 1$ (log preferences) the initial consumption response is virtually identical. When $\gamma = 0.5$ both are very small. When $\gamma = 5$ agents are less willing to substitute consumption over time and the wealth effect is stronger, but the difference between model $\mathcal{A}$ and G is still only 1 percentage point. Table 6 also shows that the responses of the short term interest rate are nearly identical under models $\mathcal{A}$ and $\mathcal{G}$, even for relatively high CRRA. The parameters of monetary policy (e.g., the natural rate) are therefore also similar. Moreover these results obtain even though agents are assumed to observe the trend directly. Adding learning and imperfect information, as in Section 1.2 or in Edge et al. (2007), would make the responses even mode similar.

The conclusion, then, is that the behavior of the economy, *conditional on the same path for A_t*, would be roughly the same under models A or G. To understand this point, note that the state space is (A_t, K_{t-1}) and that capital is determined jointly by the policy function for C and the law of motion: $K_t = \mathscr{K}(K_{t-1}, A_t; \mathcal{M})$ where the function $\mathscr{K}$ is not sensitive to the model $\mathcal{M}$ used by the agents to interpret the data. The same path of $A(t)$ therefore leads to (approximately) the same path for K_t and to the same path of C_t. Given the same historical path of TFP, the two models make similar predictions about consumption, investment, the labor share, the capital labor ratio or inflation. Recognizing that TFP growth is linear therefore does not affect existing research on these topics.[11]

The response of long term interest rates is somewhat stronger under model $\mathcal{G}$ than under model $\mathcal{A}$. For instance, when $\gamma = 2$ the growth shock increases the 20 year rate by 2.68p.p. under model $\mathcal{A}$ ($b = 0.02$) versus 3.24p.p. under model $\mathcal{G}$ ($g = 2\%$). The stronger response of the long rate is consistent with the expected path of consumption but it also explains the similarity of the initial response of consumption and investment. Agents anticipate higher returns to capital in the future under model $\mathcal{G}$ than under model $\mathcal{A}$, which, all else equal, would increase Tobin's Q and investment. The increase in the long rate compensates that difference so that the response of Tobin's Q becomes similar in both models. When $\gamma = 0.5$ Tobin's Q increases by 2.2p.p. under model $\mathcal{A}$ versus 2.5p.p. under model $\mathcal{G}$. When $\gamma = 2$, Tobin's Q decreases by 6.9p.p. under $\mathcal{A}$ versus 7.6p.p. under $\mathcal{G}$.

5.2 Endogenous Growth

Endogenous growth models were designed to generate exponential growth. I now present a simple model that delivers additive growth. I use an expanding variety model à la Romer (1990) to illustrate my point but it is relatively straightforward to apply it to the

[11] Model $\mathcal{G}$ predicts balanced growth if g happens to be constant, while model $\mathcal{A}$ does not predict balanced growth until the rate of growth is zero. However, since g is not constant in the data the fact that model $\mathcal{G}$ makes it easier to compute a balanced growth is irrelevant.

the other types of endogenous growth models (AK, human capital (Uzawa, 1965; Lucas, 1988) or quality ladders (Aghion and Howitt, 1992)). Following the textbook treatment in Barro and Sala-i-Martin (2004) the final good is produced using intermediate inputs and labor using a production function from Spence (1976)

$$Y_t = AL_t^{1-\alpha} \int_0^{N_t} x_{i,t}^{\alpha} di,$$

where N_t is the set of varieties that have been discovered up to time t and labor supply L_t is exogenous. The price of final output is normalized to one.

Micro given N_t The demand curve for product i at time t, $\alpha A L_t^{1-\alpha} x_{i,t}^{\alpha-1} = p_{i,t}$, is iso-elastic with price elasticity $\epsilon = 1/(1-\alpha)$. The wage is such that $w_t L_t = (1-\alpha) Y_t$. Variety i is produced from final output with a constant marginal cost ψ so profits are $\pi_{i,t} = (p_{i,t} - \psi) x_{i,t}$. The profit maximizing price is $p_{i,t} = \frac{\psi}{\alpha}$ and output per variety is

$$x_{i,t} = \left(\alpha^2 \frac{A}{\psi} \right)^{\frac{1}{1-\alpha}} L_t. \tag{13}$$

An important feature of (13) is that the quantity of each input is independent of N_t. It only depends on productivity and market size measured by L_t. Profits are also proportional to market size: $\pi_t = \bar{\pi} L_t$ where $\bar{\pi} = \frac{1-\alpha}{\alpha} (\alpha^2 A)^{\frac{1}{1-\alpha}} \psi^{-\frac{\alpha}{1-\alpha}}$. Compared to the first best – $p^* = \psi$ and $x_t^* = \left(\alpha \frac{A}{\psi} \right)^{\frac{1}{1-\alpha}} L_t$ – output is too low because of market power.

Macro given N_t Given N_t, production is

$$Y_t = A^{\frac{1}{1-\alpha}} \left(\frac{\alpha^2}{\psi} \right)^{\frac{\alpha}{1-\alpha}} N_t L_t, \tag{14}$$

and labor productivity $A^{\frac{1}{1-\alpha}} \left(\frac{\alpha^2}{\psi} \right)^{\frac{\alpha}{1-\alpha}} N_t$ is proportional to N_t. Output is either consumed, used to produce existing varieties, or used to generate new varieties. Market clearing thus requires $Y_t = C_t + \psi X_t + \bar{\kappa}_t \dot{N}_t$, where $\bar{\kappa}_t$ is the average cost per new variety that I discuss below. Using the equilibrium conditions we see that $C_t = (1-\alpha^2) Y_t - \bar{\kappa}_t \dot{N}_t$. From $w_t L_Y = (1-\alpha) Y_t$ we have the equilibrium wage $w_t = (1-\alpha) A^{\frac{1}{1-\alpha}} \left(\frac{\alpha^2}{\psi} \right)^{\frac{\alpha}{1-\alpha}} N_t$. Consumers have a standard Euler equation, expressed in per capita consumption $c_t = C_t/L_t$ as

$$\gamma \frac{\dot{c}_t}{c_t} = r_t - \rho. \tag{15}$$

Innovation I assume that the innovator gets a permanent patent so the value of discovering a new variety is

$$v_t = \int_t^\infty \pi_\tau e^{-(\tau - t) r_{t,\tau}} d\tau \tag{16}$$

where $r_{t,\tau}$ is the zero coupon yield at time t with maturity τ. Up to this point the model is exactly the same irrespective of the dynamics of TFP. I now propose a specification that nests the exponential model as a special case. At any time t there is a set M_t of ideas for new varieties, indexed by their discovery costs: idea j costs κ_j. The marginal idea financed at time t must satisfy $\kappa_{j(t)} = v_t$. The number of new varieties is therefore

$$\dot{N}_t = F(v_t) M(L_t, N_t) \tag{17}$$

where $F(.)$ is the c.d.f. of κ_j and $M_t = M(L_t, N_t)$ is the mass of potential ideas. A simple model could be that each person has some probability of having an idea, in which case M_t would be proportional to L_t. M_t could also depend on the number N_t of varieties already discovered. In general I write $M_t = M(L_t, N_t)$.

The dynamics of the system are pinned down by the Euler equation (15), the NPV of monopoly profits (16), and the variety production equation (17). I focus for simplicity on a balanced growth path with constant labor $L_t = L$. With constant interest rate r we have $v = \frac{\pi}{r}$. From the resource constraint we see that the growth rate of consumption per capita is the same as the growth rate of N. The dynamics are therefore

$$\dot{N}_t = F\left(\frac{\pi}{r}\right) M(L, N_t) \tag{18}$$

$$\gamma \frac{\dot{N}_t}{N_t} = r - \rho \tag{19}$$

I now explain how the function M determines the nature of growth.

Exponential Growth Let us start with the "standard" model. To get exponential growth one has to assume that the number of new ideas is *proportional* to the number of existing varieties $M_t = \bar{M} N_t$. Growth is such that $\gamma g_c = r - \rho$ and the equilibrium interest rate solves

$$\gamma \bar{M} F\left(\frac{\pi}{r}\right) = r - \rho. \tag{20}$$

The comparative statics are the same as in the textbook model: Growth is high when risk aversion is low, when consumers are patient, when innovation costs are low, and when monopoly profits are high. The model has a scale effect because π is proportional to L – larger markets enable higher profits and faster accumulation of N.

Equation (20) allows us to better understand the key assumption of the standard model. In textbooks (e.g., Barro and Sala-i-Martin, 2004) one often assumes an *infinite* supply of new ideas at constant cost κ. Technically this corresponds to $M = \infty$ together with a degenerate distribution of potential ideas: F is discontinuous at κ: $F(\kappa^-) = 0$ while $F(\kappa^+) = 1$. In that special case we must have $r = \frac{\pi}{\kappa}$ and we obtain the growth rate directly from the Euler equation $\gamma g_c = \frac{\pi}{\kappa} - \rho$. Infinite elasticity at κ is unrealistic[12] and it hides the fundamental assumption driving exponential growth, namely that M must be proportional to N. Proportionality ensures that the capacity constraint on new ideas does not become binding as the economy grows. It is all that is needed for balanced growth, it is much weaker than assuming $M = \infty$, and it allows a more transparent comparison with other models.

Remark 1. The fundamental assumption delivering exponential growth in the standard model is that the number of new ideas M_t is proportional to the number of existing varieties N_t.

All endogenous growth models make the same fundamental assumption to obtain exponential growth. The quality ladder model, for instance, assumes that the steps of the ladder are exponentially distributed. The key assumption that ideas *multiply* each other does not seem particularly plausible. In the context of expanding varieties, (14) implies that labor productivity is proportional to N, which makes sense since these varieties were invented precisely to offer goods that consumers want. But there is no reason to think that the flow of new ideas should also be proportional to N.

Additive Growth We obtain an additive growth model when M does not depend on N. In the long run the interest rate converges to the rate of time preference ρ and the

[12] It says that all varieties existing today could have been discovered earlier and the only reason they were not is that the cost in terms of consumption would have been too high. Strictly speaking, this assumption means that the iPhone could have been invented in 1900.

increment is constant: $\dot{N} = F\left(\frac{\pi}{\rho}\right) M$ implies

$$N = N_0 + t \times F\left(\frac{\pi}{\rho}\right) M. \tag{21}$$

The Lemma summarizes our discussion so far

Lemma 1. *Growth is exponential when M_t is proportional to N_t and additive when M_t does not depend on N_t.*

If we think that ideas occur in people then a natural specification is $M = \bar{M}L$. Aggregate innovation $\dot{N} = F\left(\frac{\bar{\pi}L}{\rho}\right) \bar{M}L$ depends on market size in two ways: big countries have big markets that can sustain more varieties ($\bar{\pi}L/\rho$) and they have more people who can generate ideas ($\bar{M}L$). This scale effect does not alter the fact that growth is additive. Time varying population growth, L_t, would generate time varying economic growth just as in the standard model.

Recent research has discussed the evolution of R&D spending and research productivity. Jones (2009) argues that researchers spend an increasing amount of time getting to the frontier of knowledge before they can finally push it. Bloom et al. (2020) show that the number of researchers has increased significantly over the past 80 years. Since TFP growth has not increased they conclude that research productivity has declined. This literature uses exponential growth as a benchmark therefore additive growth would appear as a form of decreasing return. Note, however, that additive growth does not mean stagnation: productivity goes to infinity, just not as quickly as in the exponential benchmark.

With decreasing returns to R&D the models cannot sustain endogenous growth, which is why Bloom et al. (2020) and Jones (2021) argue for semi-endogenous models where population growth is the ultimate source of long term growth. Population growth can overcome decreasing returns in two ways. The pull factor comes from increasing market size and profits since $\pi_t = \bar{\pi}L_t$. The push factor comes from the increase in the number of ideas as M increases with L.

The additive growth equation (21) fits the TFP data well. Its implications for R&D spending appear more ambiguous. In the very long run, assuming no break and once r has converged to ρ, we have $v_\infty = \pi/\rho$. At this point the total resources spent on innovation – $M \int_0^{v_\infty} \kappa dF(\kappa)$– is fixed and since productivity is growing linearly the share of spending in GDP falls.[13] Along this path, however, v increases as r falls and thus

[13]That it decreases towards zero is an artifact of the accounting for research effort. In the simple

spending increases and the evolution of the spending share is ambiguous. Moreover, in the data we observe important changes in the valuation equation. Assuming that new ideas are embedded in existing firms, the ratio of aggregate firm value to GDP is proportional to $\frac{\pi}{r-g}$ and this ratio has increased over time despite the fall in g, perhaps because of a decline in the discount rate (lower risk free rate or lower equity risk premium), or perhaps because firms can appropriate a higher fraction of the value of their innovations. Conditional on the observed value to GDP ratio the model would predict a stable or increasing the ratio of R&D spending to GDP.[14]

Mean-Field Models of Endogenous Growth This discussion can be extended to models of growth through learning and interactions, as in Lucas and Moll (2014), Perla and Tonetti (2014) and Akcigit et al. (2018). In these models agents actively seek meetings with others to learn productivity-increasing ideas. If the cost of searching is measured by foregone production then low productivity agents search more. As a result, the left tail of low productivity is replaced by draws from the upper tail. This is the process that generates growth. The growth rate depends on characteristics of the productivity distribution, with a thicker-tailed distribution leading to more growth. The planner's problem takes into account and internalizes the external benefits of search.

Balanced exponential growth in these models requires an unbounded Pareto distribution of initial productivities. Perla and Tonetti (2014), however, show that even if the distribution is bounded-Pareto, growth can be approximately constant for many years. Similarly, suppose that the initial distribution of productivity is uniform over $[a_0, A_{max}]$ with an average of $\bar{A}_0 = (a_0 + A_{max})/2$. Suppose that firms with productivities between a_0 and $a_0 + b$ choose to search and imitate. The distribution of productivity at time 1 will be uniform over $[a_1, A_{max}]$ with $a_1 = a_0 + b$. If the process continues, average productivity grows additively as $\bar{A}_{t+1} = \bar{A}_t + b/2$. This process cannot go on indefinitely, but, just as in Perla and Tonetti (2014), it can be a good approximation for many years when a_t/A_{max} is small.[15]

model all the costs of innovation are made when the variety is created. In practice much R&D spending is ongoing, which changes the accounting without changing the model's prediction. Suppose that the owner of a variety must spend $m < \pi$ units per period to maintain the relevance of her variety. The flow profits become $\pi - m$ and the model is unchanged by simply replacing π with $\pi - m$ in the equilibrium conditions. In that case the ratio of R&D to GDP converges to m in the additive model.

[14] Additive growth requires $F_t(v_t)$ be constant. An increase in v (say because of lower risk premia) and a decrease in F (say because ideas are more expensive) can then explain additive growth with increasing spending on innovation.

[15] One must sill compute the value function and verify that the optimal policy is to imitate when firm productivity is in $[a_t, a_t + b]$ with b (approximately) constant. Computations are available upon

5.3 Inter-temporal Spillovers

An important difference between models $\mathcal{A}$ and $\mathcal{G}$ lies in the strength of inter-temporal spillovers. Suppose that productivity A_t evolves along some equilibrium and consider a one time deviation where research output increases by ϵ/Δ from time t_0 to $t_0+\Delta$ (so ϵ is the cumulative increase). At any point $t > t_0 + \Delta$ TFP becomes $A'_t = A_t + \epsilon$ under additive growth, and $A'_t = A_t e^{\epsilon}$ under exponential growth. In other words

$$\left.\frac{\partial A_t}{\partial \epsilon_0}\right|_{\mathcal{A}} = 1$$

while

$$\left.\frac{\partial \tilde{A}_t}{\partial \epsilon_0}\right|_{\mathcal{G}} = A_t$$

In the exponential model, the impact on future productivity of a small change at t_0 becomes infinitely large as we extend the time horizon. The following thought experiment shows why this is implausible. Suppose that the US TFP shift of 1930 happened in 1910 instead. Using our estimates of a change in local growth rate from 1% to 3.3% during that window, the exponential model says that TFP today (2020) would be 58% higher and labor productivity would be twice as high. Believing in exponential growth means believing that if the US had implemented electricity 20 years earlier than it actually did, GDP today would be twice as high (holding constant the quantity of hours worked). This seems wildly implausible. The linear model, by contrast, says that TFP would be higher by 0.8 points from a baseline of 8, so 10% higher. Taking capital accumulation into account, GDP would then be about 15% higher, as opposed to 100% higher under exponential growth.

Planner's Solution A simple way to highlight the role of inter-temporal spillovers is to compare the planner allocation with the decentralized equilibrium. We already know that the static equilibrium is inefficient. The planner would set the price $p^* = \psi$, the quantity $x_t^* = \left(\frac{\alpha}{\psi}\right)^{\frac{1}{1-\alpha}} AL$ and the level of output would be $Y_t^* = \left(\frac{\alpha}{\psi}\right)^{\frac{\alpha}{1-\alpha}} ALN_t$. The resource constraint is

$$Y_t = C_t + \psi N_t x_t + M(L, N_t) \int_0^{\kappa_t} \kappa dF(\kappa)$$

request.

Given the Cobb-Douglas production function the resources spent on intermediate goods are $\psi N_t x_t = \alpha Y_t$ so the resource constraint in per- capita terms is

$$c_t = \tilde{A} N_t - \frac{M(L, N_t)}{L} \int_0^{\kappa_t} \kappa dF(\kappa)$$

where $\tilde{A} \equiv (1-\alpha) A \left(\frac{\alpha}{\psi}\right)^{\frac{\alpha}{1-\alpha}}$. Since the innovation equation is $\dot{N}_t = F(\kappa_t) M(L, N_t)$ I can write the planner's Hamiltonian as

$$\mathcal{H} = u\left(\tilde{A} N_t - \frac{M(L, N_t)}{L} \int_0^{\kappa_t} \kappa dF(\kappa)\right) e^{-\rho t} + \lambda_t F(\kappa_t) M(L, N_t),$$

where λ_t is the Lagrange multiplier and κ_t the control variable. The optimal κ_t must satisfy $\frac{\kappa_t}{L} u'(c_t) e^{-\rho t} = \lambda_t$ and the co-state equation is$\left(\tilde{A} - \frac{\partial M_t}{\partial N_t} \frac{\int_0^{\kappa_t} \kappa dF(\kappa)}{L}\right) u'(c_t) e^{-\rho t} + \lambda_t F(\kappa_t) \frac{\partial M_t}{\partial N_t} + \dot{\lambda}_t = 0$. Assuming CRRA preferences with $\gamma = -\frac{c_t u"_t}{u'_t}$ we get

$$\gamma \frac{\dot{c}_t}{c_t} = \tilde{A} \frac{L}{\kappa_t} + \frac{\dot{\kappa}_t}{\kappa_t} - \rho + \frac{\partial M_t}{\partial N_t}\left(F(\kappa_t) - \frac{1}{\kappa_t} \int_0^{\kappa_t} \kappa dF(\kappa)\right)$$

We can now compare the Planner's solution to the decentralized one. Assuming production subsidies that undo the static markup the decentralized equilibrium is

$$\gamma \frac{\dot{c}_t}{c_t} = \frac{\tilde{A} L}{v_t} + \frac{\dot{v}_t}{v_t} - \rho$$
$$\dot{N}_t = F(v_t) M(L, N_t)$$

We see that the difference between the planner and the (statically efficient) private equilibrium comes from $\frac{\partial M_t}{\partial N_t}\left(F(\kappa_t) - \frac{1}{\kappa_t} \int_0^{\kappa_t} \kappa dF(\kappa)\right)$. This term is zero in the additive model since $\frac{\partial M_t}{\partial N_t} = 0$. It is also zero in the infinite elasticity model since $F(\kappa_t) - \frac{1}{\kappa_t} \int_0^{\kappa_t} \kappa dF(\kappa) = 0$ when F is discontinuous at some fixed $\bar{\kappa}$. In all other cases $F(\kappa_t) - \frac{1}{\kappa_t} \int_0^{\kappa_t} \kappa dF(\kappa) > 0$ and the externality has the sign of $\frac{\partial M_t}{\partial N_t}$. Under exponential growth, then, the planner wants to subsidize R&D. Under additive growth the planner does not want to subsidize R&D.

5.4 A Simple Rejoinder

Section 4 shows that the trend growth of the technology frontier changes over time but these changes are not explained by an exponential model. They highlight instead the role of technological revolutions and GPTs. A simple model that describes TFP is as follows:

$$A_t - A_{t-1} = b_t + \epsilon_t,$$

where ϵ is iid. There is a small probability p of a regime change as

$$b_{t+1} = \begin{cases} b_t & , 1-p \\ A_t \xi_{t+1} & , p \end{cases}$$

I normalize the new regime by the level of TFP at the time of the regime change so that the specification nests models $\mathcal{A}$ and $\mathcal{G}$:

$$\mathbb{E}_t \left[A_{t+1} - A_t \right] = (1-p)\, b_t + p A_t \mathbb{E}_t \left[\xi_{t+1} \right]. \tag{22}$$

The pure model $\mathcal{A}$ corresponds to $p = 0$, the pure model $\mathcal{G}$ to $p = 1$. The historical data suggests $p \leq 1\%$ per annum which explains the success of model $\mathcal{A}$. Growth is linear within a GPT era, but there is small chance of discovering a new GPT. With the normalization by A_t we have $\xi_{1650} = 0.35\%$, $\xi_{1830} = 0.94\%$, and $\xi_{1930} = 4.4\%$. The structural change of the 1930s appears truly amazing in that respect.

Equation (22) is related to equation (2) in Comin et al. (2010). They point out that the exponential nature of growth depends on the complementarity between old and new technologies. My results suggest that, in most times and places new technologies increase TFP independently of existing technologies. One could interpret a GPT as a technological change that is complementary to a sufficiently high share of existing ideas and technologies. This complementarity creates what looks like multiplicative growth as the GPT is implemented. The key point is that the TFP equation changes following the discovery of a new GPT. The linear growth equation holds within each GPT era but not across GPTs. An important question for future research is the persistence of GPTs. Should we assume that a GPT permanently increases the (potential) growth of the economy? Or should we assume that the impact on b depreciates over time? One could speculate that the slowdown of the late 1970s in Figure 12 reflects the waning impact of the initial electricity revolution and the pickup in the late 1980s the impact of IT. This is an interesting question for future research.

6 Conclusion

TFP growth is not exponential. TFP has been growing linearly over the past 90 years in the US and the additive model beats the exponential model for every country – developed or developing – where TFP data is available. The TFP frontier appears to grow linearly within broad historical periods: 1650 to 1830, 1830 to 1930, and 1930 until today. Additive TFP growth predicts *increasing* growth of labor productivity and GDP per capita thanks to capital accumulation.

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. The additive model does not necessarily solve the research productivity puzzle of Bloom et al. (2020) since this puzzle is not about the stochastic process for TFP but rather about the specification of the production function for ideas.

The additive model, unlike the exponential one, provides useful long run forecasts. To illustrate this point consider the predictions one would make in 2020 regarding GDP in 2060, holding population constant so as not to introduce additional demographic forecast errors. TFP level is around 3 in 2020. The estimate for TFP growth is 1.2% with a standard deviation of 0.2% over the preceding 40 years. The estimate for TFP increments is 0.027 with a standard deviation of 0.0036. The $\mathcal{G}$-forecast for cumulative growth between 2020 and 2060 is 2 (i.e., $1.012^{\frac{40}{1-\alpha}}$) but the two standard errors range is 1.6 to 2.6, which is \$21 trillion. It is difficult to see the usefulness of a forecast where the error range is as large as the current value of GDP. The $\mathcal{A}$-forecast is 1.59 with a range of 1.42 to 1.76, which is only one third of 2020 GDP.

Additive growth has implications for macroeconomic, industry and firms dynamics. At a theoretical model, additive growth suggests that new ideas add to our stock of knowledge but do not multiply it. Per capita income and consumption growth is lower and less volatile under additive growth than under exponential growth. This affects the valuation of long term assets (e.g. pensions) and the quantity of long term risk in the economy. This matters for the optimal mitigation of long term risks such as climate change, since real discount rates are low and future generations will not be much richer than the current one. At the firm level, Lenzu et al. (2023) show 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.

References

Aghion, P. and P. Howitt (1992): "A Model of Growth Through Creative Destruction," *Econometrica*, 60, 323–351.

Akcigit, U., S. Caicedo, E. Miguelez, S. Stantcheva, and V. Sterzi (2018): "Dancing with the Stars: Innovation through Interactions," NBER WP 24466.

Andrew, G., J. B. Carlin, H. S. Stern, A. Vehtari, and D. B. Rubin (2021): *Bayesian Data Analysis*.

Bai, J. and P. Perron (2003): "Computation and analysis of multiple structural change models," *Journal of Applied Econometrics*, 18, 1–22.

Barro, R. J. and X. Sala-i-Martin (2004): *Economic Growth*, Cambridge, Massachusetts: MIT Press.

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.

Bolt, J. and J. L. van Zanden (2020): "Maddison style estimates of the evolution of the world economy. A new 2020 update," Maddison Project Database.

Bouscasse, P., E. Nakamura, and J. Steinsson (2021): "When Did Growth Begin? New Estimates of Productivity Growth in England from 1250 to 1870," Working Paper.

Bresnahan, T. and M. Trajtenberg (1995): "General purpose technologies: Engines of growth?" *Journal of Econometrics*, 65, 83–108.

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

Comin, D. and B. Hobijn (2010): "An Exploration of Technology Diffusion," *American Economic Review*, 100, 2031–59.

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

Durbin, J. and S. J. Koopman (2008): "Time Series Analysis by State Space Methods," *FRB of New York Staff Report*.

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

Fernald, J. G. (2012): "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," FRBSF Working Paper (updated March 2014).

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

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

Grossman, G. M. and E. Helpman (1991): "Quality Ladders in the Theory of Growth," *The Review of Economic Studies*, 58, 43–61.

Groth, C., K.-J. Koch, and T. Steger (2010): "When economic growth is less than exponential," *Economic Theory*, 44, 213–242.

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.

——— (2002): "Sources of U.S. Economic Growth in a World of Ideas," *American Economic Review*, 92, 220–239.

——— (2021): "The Past and Future of Economic Growth: A Semi-Endogenous Perspective," Annual Review of Economics.

Jovanovic, B. and P. L. Rousseau (2005): "General Purpose Technologies," in *Handbook of Economic Growth*, ed. by P. Aghion and S. Durlauf, Elsevier, chapter 18.

Kruse-Andersen, P. (2022): "Testing RD-Based Endogenous Growth Models," Working paper.

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

Ljungqvist, L. and T. J. Sargent (2018): *Recursive Macroeconomic Theory*, MIT Press.

Lucas, R. E. (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22, 3–42.

Lucas, R. E. and B. Moll (2014): "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 122, 1–51.

Mokyr, J. and H.-J. Voth (2010): "Understanding growth in Europe, 1700-1870: theory and evidence," in *The Cambridge Economic History of Modern Europe*, ed. by S. Broadberry and K. H. O'Rourke, Cambridge University Press, vol. 1.

Perla, J. and C. Tonetti (2014): "Equilibrium Imitation and Growth," *Journal of Political Economy*, 122, 52–76.

Romer, P. M. (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98, S71–S102.

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.

Spence, M. (1976): "Product Selection, Fixed Costs, and Monopolistic Competition," *The Review of Economic Studies*, 43, 217–235.

Uzawa, H. (1965): "Optimum Technical Change in An Aggregative Model of Economic Growth," *International Economic Review*, 6, 18–31.

Voth, H.-J. (2001): *Time and Work in England, 1750-1830*, Oxford University Press.

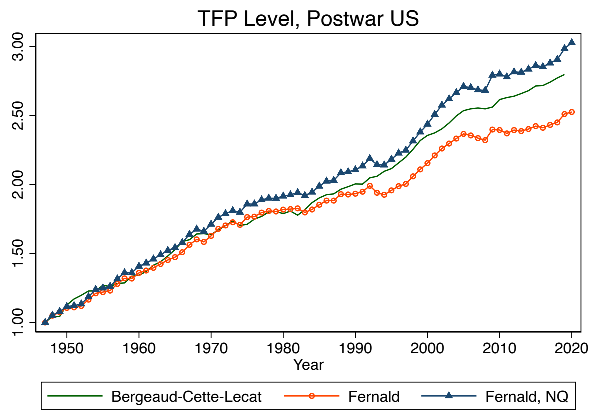
Appendix

A US Data

A.1 Three Measures of Post-War US TFP

Figure 15 compares the TFP series from BCL and Fernald, with and without adjustment for education.

Figure 15: US TFP Levels



Notes: TFP levels, A^{bcl}, A_t^f, and A_t^n. Data from Fernald (2012) and Bergeaud et al. (2016).

A.2 Labor Productivity

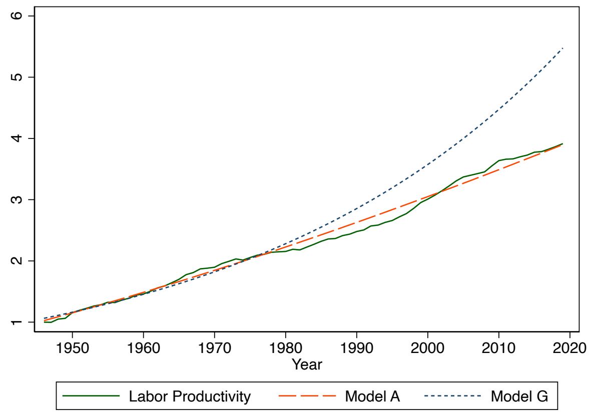
Let us now study the accumulation of capital. Define the capital labor ratio as

$$k_t \equiv K_t/L_t,$$

where, in the BCL data, K_t is the real capital stock and L_t measures hours worked. The first order condition for capital demand in the neoclassical growth model equates the marginal product of capital (MPK) to the user cost (defined as χ). BCL do not consider changes in the user cost and the first order condition is simply

$$k_t^{1-\alpha} = \frac{\alpha}{\chi} A_t. \tag{23}$$

Figure 16: Out-of-Sample Labor Productivity Forecasts



Notes: Models are estimated over 1947-1983. The forecast 1984-2019 is out-of-sample. Labor productivity is real GDP per hour. Data source: Bergeaud et al. (2016).

Equation (23) says that the normalized inverse MPK (IMPK) is proportional to A.[16] Model $\mathcal{G}$ therefore predicts that $k_t^{1-\alpha}$ grows exponentially, while model $\mathcal{A}$ says that it grows linearly. Once we have a forecast for the capital labor ratio we can use our forecast for TFP to create a forecast for labor productivity λ_t, defined as output per hour:

$$\lambda_t \equiv \frac{Y_t}{L_t} = A_t k_t^{\alpha}. \tag{24}$$

model $\mathcal{A}$ offers a forecast for labor productivity as

$$\hat{\lambda}_t = \left(\hat{a} + \hat{b}t\right)\left(\hat{a}_{impk} + \hat{b}_{impk}t\right)^{\frac{\alpha}{1-\alpha}}$$

B Bayesian Model Selection

B.1 Model $\mathcal{A}$

Consider the corresponding steady-state innovations representation (Ljungqvist and Sargent, 2018) of model $\mathcal{A}$:

[16] Users of model $\mathcal{G}$ typically interpret equation (23) as saying that capital grows exponentially, just like A, as a rate $(1+g)^{1/(1-\alpha)}$. Equivalently, if the model is written with Harrod-neutral technological progress, $Y_t = K_t^{\alpha}\left(\mathcal{M}_t H_t\right)^{1-\alpha}$ then capital is proportional to $\mathcal{M}_t$. I return to these issues in Section 5.

$$\hat{b}_t = \hat{b}_{t-1} + K a_t$$
$$A_t = A_{t-1} + \hat{b}_{t-1} + a_t$$

where $\hat{b}_t = \mathbb{E}\left[b_{t+1}|A^t\right]$, $\mathbb{E}\left[a_t^2\right] = \Omega$ for all $t \geq 0$, and K is the steady-state Kalman gain. First, let us consider A_{t+1} :

$$A_{t+1} = A_t + \hat{b}_t + a_{t+1}$$

Define $A_{t+h|t}$ as the random variable A_{t+h} conditional on time-t information. The moments of the 1-step ahead prediction errors are then given by:

$$\mathbb{E}^{\mathcal{A}}\left[A_{t+1|t}\right] = A_t + \hat{b}_t$$
$$\mathtt{Var}^{\mathcal{A}}\left[A_{t+1|t}\right] = \Omega$$

This last equation is about the covariance of forecasts made at different times for a fixed horizon ($h = 1$ here). Note that $\mathtt{Cov}^{\mathcal{A}}\left[A_{t+1|t}, A_{t+2|t+1}\right] = \mathbb{E}^{\mathcal{A}}\left[a_{t+1} * a_{t+2}\right] = 0$. Sequential one-period ahead forecast errors are uncorrelated.

Consider next the two-period forecasts ($h = 2$). We have

$$\begin{aligned} A_{t+2} &= A_{t+1} + \hat{b}_{t+1} + a_{t+2} \\ &= A_t + \hat{b}_t + a_{t+1} + \hat{b}_t + K a_{t+1} + a_{t+2} \\ &= A_t + 2\hat{b}_t + (1+K)\, a_{t+1} + a_{t+2} \end{aligned}$$

The moments are therefore

$$\begin{aligned} \mathbb{E}^{\mathcal{A}}\left[A_{t+2|t}\right] &= A_t + 2\hat{b}_t \\ \mathtt{Var}^{\mathcal{A}}\left[A_{t+2|t}\right] &= \left((1+K)^2 + 1\right)\Omega \\ \mathtt{Cov}^{\mathcal{A}}\left[A_{t+2|t}, A_{t+1|t-1}\right] &= (1+K)\,\Omega \end{aligned}$$

The forecast errors are now correlated because the intervals overlap. If a_{t+1} is high, the forecast for $A_{t+1|t-1}$ is too low and the forecast for $A_{t+2|t}$ is also likely to be too low.

When $h = 3$ the overlap is longer and two-sided

$$A_{t+3|t} = A_t + 3\hat{b}_t + (1+2K)\, a_{t+1} + (1+K)\, a_{t+2} + a_{t+3}$$

The moments are therefore

$$\begin{aligned}
\mathbb{E}^{\mathcal{A}}\left[A_{t+3|t}\right] &= A_t + 3\hat{b}_t \\
\mathtt{Var}^{\mathcal{A}}\left[A_{t+3|t}\right] &= \left(1 + (1+K)^{\,2} + (1+2K)^2\right)\Omega
\end{aligned}$$

and the covariances are

$$\begin{aligned}
\mathtt{Cov}^{\mathcal{A}}\left[A_{t+3|t}, A_{t+2|t-1}\right] = \mathtt{Cov}^{\mathcal{A}}\left[A_{t+3|t}, A_{t+4|t+1}\right] &= \left[(1+K)(1+2K) + 1 + K\right]\Omega \\
\mathtt{Cov}^{\mathcal{A}}\left[A_{t+3|t}, A_{t+1|t-2}\right] = \mathtt{Cov}^{\mathcal{A}}\left[A_{t+3|t}, A_{t+5|t+2}\right] &= (1+2K)\,\Omega
\end{aligned}$$

and $\mathtt{Cov}^{\mathcal{A}}\left[A_{t+3|t}, A_{t+3+p|t+p}\right] = 0$ for all $|p| > 2$.

We can generalize these moments as follows:

$$\begin{aligned}
\mathbb{E}^{\mathcal{A}}\left[A_{t+h|t}\right] &= A_t + h\hat{b}_t \\
\mathtt{Var}^{\mathcal{A}}\left[A_{t+h|t}\right] \equiv \chi^{\mathcal{A}}_{0,h} &= \Omega \sum_{\tau=0}^{h-1} (1+\tau K)^2 \\
\mathtt{Cov}^{\mathcal{A}}\left[A_{t+h,|t}, A_{t+h+p|t+p}\right] \equiv \chi^{\mathcal{A}}_{p,h} &= \begin{cases} \Omega \sum_{\tau=1}^{h-|p|} \left(1 + (h-\tau)\,K\right)\left(1 + (h-\tau-|p|)\,K\right) & \text{for } |p| < h \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

where the relevant range for time series index p is $[-h+1; h-1]$.

B.2 Model $\mathcal{G}$

Model $\mathcal{G}$ is more complicated because of the multiplicative growth terms. In this case, I approximate the moments of the forecast distribution via Monte Carlo simulations. I use 8,000 simulations to perform the calculations, and confirm that the results are robust to this choice.[17] These simulations give approximations for $\mathbb{E}^{\mathcal{G}}\left[A_{t+h}\right]$, $\mathtt{Var}^{\mathcal{G}}\left[A_{t+h}\right] \equiv \chi^{\mathcal{G}}_{0,h}$ and $\mathtt{Cov}^{\mathcal{G}}_t\left[A_{t+h}, A_{t+h+p}\right] \equiv \chi^{\mathcal{G}}_{p,h}$. Under our simplification that $\hat{g}_t$ is known with certainty, these moments completely characterize the distribution of the h-period ahead forecasts

[17]To further confirm the validity of the Monte Carlo approximation, I also do the MC approximations for Model $\mathcal{A}$, and confirm that the approximated covariance matrices are close in supremum norm to the those analytically derived above.

under model $\mathcal{G}$.

Because the fixed-horizon forecasts are serially correlated over time, we must calculate the likelihood of the vector of forecasts in its entirety.[18] In other words, given a forecast horizon h, the vectors $A_{1+h|1}, A_{2+h|2}, \ldots, A_{T|T-h}$, under either model $\mathcal{A}$ or $\mathcal{G}$, are distributed as

$$\begin{bmatrix} A_{1+h|1} \\ A_{2+h|2} \\ \vdots \\ A_{T|T-h} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} A_1 + h\hat{b_2} \\ A_2 + h\hat{b_3} \\ \vdots \\ A_{T-h} + h\hat{b}_{T-h+1} \end{bmatrix}, \mathbf{V} \right) \tag{25}$$

where $\mathbf{V}$ is a matrix with χ_0^h on the diagonal, and χ_1^h on the first off-diagonal, χ_2^h on the second off-diagonal, and so on until χ_{h-1}^h for the $(h-1)$th off-diagonal. From this point, calculating the likelihood for the h-period ahead forecasts $f^h\left(A^t|\mathcal{A}\right)$ and $f^h\left(A^t|\mathcal{G}\right)$ is standard.

C Neoclassical Model

In the theoretical discussion I use continuous time and I assume an inelastic labor supply to simplify the notations. For the simulations I use a standard discrete time model with elastic labor supply. Aggregate value value added (GDP, Y_t) is given by

$$Y_t = F\left(K_t, A_t L_t\right) \tag{26}$$

where A_t is labor augmenting (Hicks-neutral) productivity,[19] L_t is the flow of labor services and K_t is the flow of capital services which accumulates as

$$\dot{K}_t = I_t - \delta K_t. \tag{27}$$

Labor grows at the constant population growth rate g_L: $\frac{dL_t}{dt} = g_L L_t$. Define $c_t = C_t/L_t$ as per capita consumption. The representative household seeks to maximize overall

[18] If it wasn't for this, we could build up the likelihood from the prediction error decomposition as standard.

[19] Note that the empirical analysis in Fernald (2012) and Bergeaud et al. (2016) uses standard growth accounting notations where $Y_t = A_t^S K_t^{\alpha} L_t^{1-\alpha}$ where A_t^S is the Solow residual. In the Cobb-Douglass case we have of course the equivalence $A_t^S = A_t^{1-\alpha}$. I have shown that A_t^S is linear in the US so strictly speaking $A_t = \left(A_t^S\right)^{\frac{1}{1-\alpha}}$ is convex.

utility

$$U = \int_0^\infty u\left(c_t\right) e^{(g_L-\rho)t} dt \tag{28}$$

where ρ is the rate of time preference and the utility function u is increasing, concave and satisfies Inada conditions. The budget constraint of households is $\frac{d\mathscr{A}_t}{dt} = r_t\mathscr{A}_t + w_t L_t - C_t$. On a per capita basis, with $a_t \equiv \mathscr{A}_t/L_t$, I obtain

$$\dot{a}_t = r_t a_t + w_t - c_t - g_L a_t.$$

Finally I rule out Ponzi schemes by imposing the condition $\lim_{t\to\infty}\left\{a_t e^{-\int_o^t (r_t - g_L)dt}\right\} \geq 0$. Assuming CRRA preferences with relative risk aversion γ we have $\frac{\dot{u'}_t}{u'_t} = -\gamma\frac{\dot{c}_t}{c_t}$ so the Euler equation with per capita consumption is

$$\gamma\frac{\dot{c}_t}{c_t} = r_t - \rho. \tag{29}$$

Population growth g_L does not appear in the Euler equation because adding a family member increases the value of higher per capita consumption in proportion to the cost of providing this extra consumption to all household members. In a closed economy households must hold the capital stock: $a_t = k_t$. I define capital per efficiency unit of labor as $\hat{k}_t \equiv \frac{K_t}{A_t L_t}$ and firms' optimal demand for capital requires

$$f'\left(\hat{k}_t\right) = r_t + \delta.$$

Finally, defining the normalized consumption as $\hat{c}_t \equiv \frac{C_t}{L_t A_t}$ I characterize the equilibrium with two equations, the capital accumulation equation

$$\dot{\hat{k_t}} = f\left(\hat{k}_t\right) - \hat{c}_t - \left(\delta + g_L + \frac{\dot{A}_t}{A_t}\right)\hat{k}_t, \tag{30}$$

and the Euler equation written in efficiency units of labor

$$\gamma\frac{\dot{\hat{c}}_t}{\hat{c}_t} = f'\left(\hat{k}_t\right) - \delta - \rho - \gamma\frac{\dot{A}_t}{A_t}. \tag{31}$$

Define

$$g_{A\infty} = \lim_{t\to\infty}\frac{\dot{A}_t}{A_t}$$

The long-term balanced growth path is given by

$$f'\left(\hat{k}_\infty\right) = \delta + \rho + \gamma g_{A\infty}$$

and

$$\hat{c}_\infty = f\left(\hat{k}_\infty\right) - \left(\delta + g_L + g_{A\infty}\right)\hat{k}_\infty$$

All per capital variables grow with A_t. For instance, long run per capita consumption is $c_t = C_t/L_t = \hat{c}_\infty A_t$. Under additive growth, the model features decreasing growth rates therefore, assuming CRRA preferences, the risk free rate falls over time and eventually converges to ρ. I can finally check the transversality condition $\lim_{t\to\infty}\left\{k_t e^{-\int_o^t (r_t - g_L)dt}\right\} = 0$. I know that $\lim_{t\to\infty}\frac{\dot{k}_t}{k_t} = g_A$ and that the long run interest rate is $r_\infty = \rho + \gamma g_A$ so the condition is

$$\rho + \gamma g_{A\infty} > g_L + g_{A\infty} \tag{32}$$

The condition says that households' discount rate must be high enough, otherwise (28) yields a value of infinity.