Tracking the Slowdown in Long-Run GDP Growth*

Juan Antolin-Diaz
Fulcrum Asset Management

Thomas Drechsel
LSE, CFM

Ivan Petrella
Warwick Business School & CEPR

Abstract

Using a dynamic factor model that allows for changes in both the long-run growth rate of output and the volatility of business cycles, we document a significant decline in long-run output growth in the United States. Our evidence supports the view that most of this slowdown occurred prior to the Great Recession. We show how to use the model to decompose changes in long-run growth into its underlying drivers. At low frequencies, a decline in the growth rate of labor productivity appears to be behind the recent slowdown in GDP growth for both the US and other advanced economies. When applied to real-time data, the proposed model is capable of detecting shifts in long-run growth in a timely and reliable manner.

Keywords: Long-run growth; Business cycles; Productivity; Dynamic factor models; Real-time data.

JEL Classification Numbers: E32, E23, O47, C32, E01.

*Antolin-Diaz: Department of Macroeconomic Research, Fulcrum Asset Management, London W1H 5BT, UK; E-Mail: juan.antolin-diaz@fulcrumasset.com. Drechsel: Department of Economics and Centre for Macroeconomics, LSE, London, WC2A 2AE, UK; E-Mail: t.a.drechsel@lse.ac.uk. Petrella: Warwick Business School, University of Warwick, Coventry, CV4 7AL, UK. E-Mail: Ivan.Petrella@wbs.ac.uk. The views expressed in this paper are solely the responsibility of the authors and cannot be taken to represent those of Fulcrum Asset Management. We thank Neele Balke, James Bell, Gavyn Davies, Davide Delle Monache, Wouter den Haan, Robert Gordon, Andrew Harvey, Ed Hoyle, Dimitris Korobilis, Juan Rubio-Ramirez, Ron Smith, James Stock, Paolo Surico, Jon Temple, Silvana Tenreyro, Joaquin Vespignani and the seminar participants at Banca d'Italia, Bank of England, Birkbeck Centre for Applied Macro, ECARES, European Central Bank, BBVA Research, the Applied Econometrics Workshop at the Federal Reserve Bank of St. Louis, the World Congress of the Econometric Society, Norges Bank and LSE for useful comments and suggestions. Alberto D'Onofrio provided excellent research assistance.

1 Introduction

"The global recovery has been disappointing (...) Year after year we have had to explain from mid-year on why the global growth rate has been lower than predicted as little as two quarters back". Stanley Fischer, August 2014.

The slow pace of the recovery from the Great Recession of 2007-2009 has prompted questions about whether the long-run growth rate of GDP in advanced economies is lower now than it has been on average over the past decades (see e.g. Fernald, 2014, Gordon, 2014b, Summers, 2014). Indeed, forecasts of US and global real GDP growth have been persistently too optimistic for the last six years.[1] As emphasized by Orphanides (2003), real-time misperceptions about the long-run growth of the economy can play a large role in monetary policy mistakes. Moreover, small changes in assumptions about the long-run growth rate of output can have large implications on fiscal sustainability calculations (Auerbach, 2011). This calls for a framework that takes the uncertainty about long-run growth seriously and can inform decision-making in real time. In this paper, we present a dynamic factor model (DFM) which allows for gradual changes in the mean and the variance of real output growth. By incorporating a broad panel of economic activity indicators, DFMs are capable of precisely estimating the cyclical comovement in macroeconomic data in a real-time setting. Our model exploits this to track changes in the long-run growth rate of real GDP in a timely and reliable manner, separating them from their cyclical counterpart.[2]

The evidence of a decline in long-run US growth is accumulating, as documented by the recent growth literature such as Fernald and Jones (2014). Lawrence Summers and Robert Gordon have articulated a particularly pessimistic view of long-run growth which contrasts

[1]For instance, Federal Open Market Committee (FOMC) projections since 2009 expected US growth to accelerate substantially, only to downgrade the forecast back to 2% throughout the course of the subsequent year. An analysis of forecasts produced by international organizations and private sector economists reveals the same pattern, see Pain et al. (2014) for a retrospective.

[2]Throughout this paper, our concept of the long run refers to changes in growth that are permanent in nature, i.e. do not mean-revert, as in Beveridge and Nelson (1981). In practice this should be thought of as frequencies lower than the business cycle.

with the optimism prevailing before the Great Recession (see Jorgenson et al., 2006). To complement this evidence, we start our analysis by presenting the results of two popular structural break tests proposed by Nyblom (1989) and Bai and Perron (1998). Both suggest that a possible shift in the mean of US real GDP growth exists, the latter approach suggesting that a break probably occurred in the early part of the 2000's.[3] However, sequential testing using real-time data reveals that the break would not have been detected at conventional significance levels until as late as mid-2012, highlighting the problems of conventional break tests for real-time analysis (see also Benati, 2007). To address this issue, we introduce two novel features into an otherwise standard DFM of real activity data. First, we allow the mean of real GDP growth, and possibly other series, to drift gradually over time. As emphasized by Cogley (2005), if the long-run output growth rate is not constant, it is optimal to give more weight to recent data when estimating its current state. By taking a Bayesian approach, we can combine our prior beliefs about the rate at which the past information should be discounted with the information contained in the data. We also characterize the uncertainty around estimates of long-run growth taking into account both filtering and parameter uncertainty. Second, we allow for stochastic volatility (SV) in the innovations to both factors and idiosyncratic components. Given our interest in studying the entire postwar period, the inclusion of SV is essential to capture the substantial changes in the volatility of output that have taken place in this sample, such as the "Great Moderation" first reported by Kim and Nelson (1999a) and McConnell and Perez-Quiros (2000), as well as the cyclicality of macroeconomic volatility as documented by Jurado et al. (2014).

When applied to US data, our model concludes that long-run GDP growth declined meaningfully during the 2000's and currently stands at about 2%, more than one percentage point lower than the postwar average. The results are supportive of a gradual decline rather than a discrete break. Since in-sample results obtained with revised data often underestimate the uncertainty faced by policymakers in real time, we repeat the exercise using real-time

[3]This finding is consistent with the analysis of US real GDP by Luo and Startz (2014), as well as Fernald (2014), who applies the Bai and Perron (1998) test to US labor productivity.

vintages of data. The model detects the fall from the beginning of the 2000's onwards, and by the summer of 2010 it reaches the significant conclusion that a decline in long-run growth is behind the slow recovery, well before the structural break tests become conclusive.

We also investigate the performance of the model in "nowcasting" short-term developments in GDP. Since the seminal contributions of Evans (2005) and Giannone et al. (2008) DFMs have become the standard tool for this purpose.[4] Interestingly, our analysis shows that standard DFM forecasts revert very quickly to the unconditional mean of GDP, so taking into account the variation in long-run GDP growth substantially improves point and density GDP forecasts even at very short horizons.

Finally, we extend our model in order to disentangle the drivers of secular fluctuations of GDP growth. Edge et al. (2007) emphasize the relevance as well as the difficulty of tracking permanent shifts in productivity growth in real time. In our framework, long-run output growth can be decomposed into labor productivity and labor input trends. The results of this decomposition exercise point to a slowdown in labor productivity as the main driver of recent weakness in GDP growth. Applying the model to other advanced economies, we provide evidence that the weakening in labor productivity appears to be a global phenomenon.

Our work is closely related to two strands of literature. The first one encompasses papers that allow for structural changes within the DFM framework. Del Negro and Otrok (2008) model time variation in factor loadings and volatilities, while Marcellino et al. (2014) show that the addition of SV improves the performance of the model for short-term forecasting of euro area GDP.[5] Acknowledging the importance of allowing for time-variation in the means of the variables, Stock and Watson (2012) pre-filter their data set in order to remove any low-frequency trends from the resulting growth rates using a biweight local mean. In his comment to their paper, Sims (2012) suggests to explicitly model, rather than filter out, these long-run trends, and emphasizes the importance of evolving volatilities for describing

[4] An extensive survey of the nowcasting literature is provided by Banbura et al. (2012), who also demonstrate, in a real-time context, the good out-of-sample performance of DFM nowcasts.

[5] While the model of Del Negro and Otrok (2008) includes time-varying factor loadings, the means of the observable variables are still treated as constant.

and understanding macroeconomic data. We see the present paper as extending the DFM literature, and in particular its application to tracking GDP, in the direction suggested by Chris Sims. The second strand of related literature takes a similar approach to decomposing long-run GDP growth into its drivers, in particular Gordon (2010, 2014a) and Reifschneider et al. (2013). Relative to these studies, we emphasize the importance of using a broader information set, as well as a Bayesian approach, which allows to use priors to inform the estimate of long-run growth, and to characterize the uncertainty around the estimate stemming both from filtering and parameter uncertainty.

The remainder of this paper is organized as follows. Section 2 presents preliminary evidence of a slowdown in long-run US GDP growth. Section 3 discusses the implications of time-varying long-run output growth and volatility for DFMs and presents our model. Section 4 applies the model to US data and documents the decline in long-run growth. The implications for tracking GDP in real time as well as the key advantages of our methodology are discussed. Section 5 decomposes the changes in long-run output growth into its underlying drivers. Section 6 concludes.

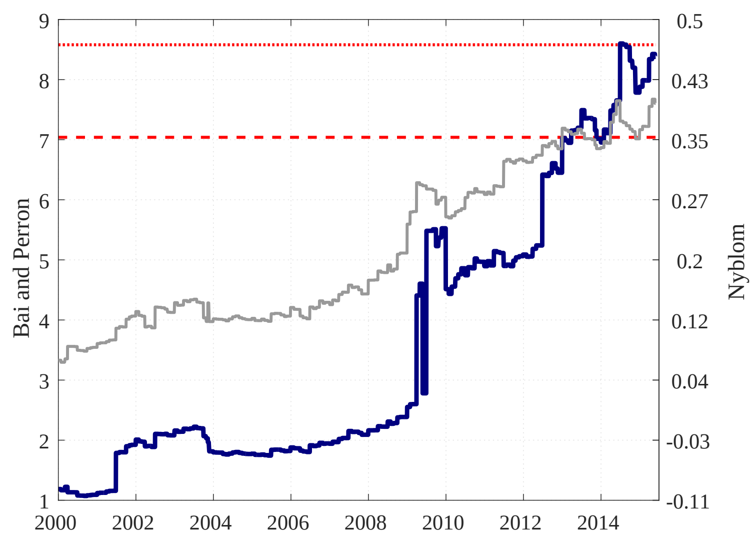
2 Preliminary Evidence

The literature on economic growth favors a view of the long-run growth rate as a process that evolves over time. It is by now widely accepted that a slowdown in productivity and long-run output growth occurred in the early 1970's, and that accelerating productivity in the IT sector led to a boom in the late 1990's.[6] In contrast, in the context of econometric modeling the possibility that long-run growth is time-varying is the source of a long-standing controversy. In their seminal contribution, Nelson and Plosser (1982) model the (log) level of real GDP as a random walk with drift. This implies that after first-differencing, the resulting growth rate fluctuates around a constant mean, an assumption still embedded in many

[6]For a retrospective on the productivity slowdown, see Nordhaus (2004). Oliner and Sichel (2000) provide evidence on the role of the IT sector in the acceleration of the late 1990's.

econometric models. After the slowdown in productivity became apparent in the 1970's, many researchers such as Clark (1987) modeled the drift term as an additional random walk, implying that the level of GDP is integrated of order two. The latter assumption would also be consistent with the local linear trend model of Harvey (1985), the Hodrick and Prescott (1997) filter, and Stock and Watson (2012)'s practice of removing a local biweight mean from the growth rates before estimating a DFM. The $I(2)$ assumption is nevertheless controversial since it implies that the growth rate of output can drift without bound. Consequently, papers such as Perron and Wada (2009), have modeled the growth rate of GDP as stationary around a trend with one large break around 1973.

Figure 1: Real-Time Test Statistics of the Nyblom and Bai-Perron Tests



Note: The gray and black solid lines are the values of the test statistics obtained from sequentially re-applying the Nyblom (1989) and Bai and Perron (1998) tests in real time as new National Accounts vintages are being published. In both cases, the sample starts in 1947:Q2 and the test is re-applied for every new data release occurring after the beginning of 2000. The dotted and dashed horizontal lines represent the 5% and 10% critical values corresponding to the two tests.

Ever since the Great Recession of 2007-2009 US real GDP has grown well below its postwar average, once again raising the question whether its mean may have declined. There are two popular strategies that could be followed from a frequentist perspective to detect parameter instability or the presence of breaks in the mean growth rate. The first one is Nyblom's (1989) L-test as described in Hansen (1992), which tests the null hypothesis of constant parameters against the alternative that the parameters follow a martingale. Modeling real GDP growth as an AR(1) over the sample 1947-2015 this test rejects the stability of the constant term at the 10% significance level.[7] The second commonly used approach, which can determine the number and timing of multiple discrete breaks, is the Bai and Perron (1998) test. This test finds evidence in favor of a single break in the mean of US real GDP growth at the 10%-level. The most likely break date is in the second quarter of 2000. In related research, Fernald (2014) provides evidence for breaks in labor productivity in 1973:Q2, 1995:Q3, and 2003:Q1, and links the latter two to developments in the IT sector. From a Bayesian perspective, Luo and Startz (2014) calculate the posterior probability of a single break and find the most likely break date to be 2006:Q1 for the full postwar sample and 1973:Q1 for a sample excluding the 2000's.

The above results indicate that substantial evidence for a recent change in the mean of US GDP growth has built up. However, the strategy of applying conventional tests and introducing deterministic breaks into econometric models is not satisfactory for the purposes of real-time decision making. In fact, the detection of change in the mean of GDP growth can arrive with substantial delay. To demonstrate this, a sequential application of the Nyblom (1989) and Bai and Perron (1998) tests using real-time data is presented in Figure 1. The evolution of the test statistics in real-time reveals that a break would not have been detected at the 10% significance levels until as late as mid-2012, which is more than ten years later than the actual break date suggested by the Bai and Perron (1998) procedure. The Nyblom (1989) test, which is designed to detect gradual change rather than a discrete break, becomes significant roughly at the same time. This lack of timeliness highlights the importance of an econometric framework capable of quickly adapting to changes in long-run growth as new information arrives.

[7] The same result holds for an AR(2) specification. In both cases, stability of the autoregressive coefficients cannot be rejected, whereas stability of the variance is rejected at the 1%-level. Section B of the Online Appendix provides the full results of both tests applied in this section.

3 Econometric Framework

DFMs in the spirit of Geweke (1977), Stock and Watson (2002) and Forni et al. (2009) capture the idea that a small number of unobserved factors drives the comovement of a possibly large number of macroeconomic time series, each of which may be contaminated by measurement error or other sources of idiosyncratic variation. Their theoretical appeal (see e.g. Sargent and Sims, 1977 or Giannone et al., 2006), as well as their ability to parsimoniously model large data sets, have made them a workhorse of empirical macroeconomics. Giannone et al. (2008) and Banbura et al. (2012) have pioneered the use of DFMs to produce current-quarter forecasts ("nowcasts") of GDP growth by exploiting more timely monthly indicators and the factor structure of the data. Given the widespread use of DFMs to track GDP in real time, this paper aims to make these models robust to changes in long-run growth. We do so by introducing two novel features into the DFM framework. First, we allow the long-run growth rate of real GDP, and possibly other series, to vary over time. Second, we allow for stochastic volatility (SV) in the innovations to both factors and idiosyncratic components, given our interest in studying the entire postwar period for which drastic changes in volatility have been documented. With these changes, the DFM proves to be a powerful tool to detect changes in long-run growth. The information contained in a broad panel of activity indicators facilitates the timely decomposition of real GDP growth into persistent long-run movements, cyclical fluctuations and short-lived noise.

3.1 The Model

Let $\mathbf{y}_t$ be an $n \times 1$ vector of observable macroeconomic time series, and let $\mathbf{f}_t$ denote a $k \times 1$ vector of latent common factors. It is assumed that $n >> k$, i.e. the number of observables is much larger than the number of factors. Formally,

$$\mathbf{y}_t = \mathbf{c}_t + \mathbf{\Lambda} \mathbf{f}_t + \mathbf{u}_t, \tag{1}$$

where $\boldsymbol{\Lambda}$ contains the loadings on the common factors and $\mathbf{u}_t$ is a vector of idiosyncratic components.[8] Shifts in the long-run mean of $\mathbf{y}_t$ are captured by time-variation in $\mathbf{c}_t$. In principle one could allow time-varying intercepts in all or a subset of the variables in the system. Moreover, time variation in a given series could be shared by other series. $\mathbf{c}_t$ is therefore flexibly specified as

$$\mathbf{c}_t = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{c} \end{bmatrix} \begin{bmatrix} \mathbf{a}_t \\ 1 \end{bmatrix}, \tag{2}$$

where $\mathbf{a}_t$ is an $r \times 1$ vector of time-varying means, $\mathbf{B}$ is an $m \times r$ matrix which governs how the time-variation affects the corresponding observables, and $\mathbf{c}$ is an $(n-m) \times 1$ vector of constants. In our baseline specification, $\mathbf{a}_t$ will be a scalar capturing time-variation in long-run real GDP growth, which is shared by real consumption growth, so that $r = 1, m = 2$. A detailed discussion of this and additional specifications of $\mathbf{c}_t$ will be provided in Section 3.2.

Throughout the paper, we focus on the case of a single dynamic factor by setting $k = 1$ (i.e. $\mathbf{f}_t = f_t$).[9] The laws of motion of the latent factor and the idiosyncratic components are

$$(1 - \phi(L))f_t = \sigma_{\varepsilon_t}\varepsilon_t, \tag{3}$$

$$(1 - \rho_i(L))u_{i,t} = \sigma_{\eta_{i,t}}\eta_{i,t}, \quad i = 1, \ldots, n \tag{4}$$

where $\phi(L)$ and $\rho_i(L)$ denote polynomials in the lag operator of order p and q, respectively. The idiosyncratic components are cross-sectionally orthogonal and are assumed to be uncorrelated with the common factor at all leads and lags, i.e. $\varepsilon_t \overset{iid}{\sim} N(0,1)$ and $\eta_{i,t} \overset{iid}{\sim} N(0,1)$.

Finally, the dynamics of the model's time-varying parameters are specified to follow

[8] The model can be easily extended to include lags of the factor in the measurement equation. In the latter case, it is sensible to avoid overfitting by choosing priors that shrink the additional lag coefficients towards zero (see D'Agostino et al., 2015, and Luciani and Ricci, 2014). We consider this possibility when we explore robustness of our results to using larger data panels in Section 4.6.

[9] For the purpose of tracking real GDP with a large number of closely related activity indicators, the use of one factor is appropriate, which is explained in more detail in Sections 4.1 and 4.2. Also note that we order real GDP growth as the first element of $\mathbf{y}_t$, and normalize the loading for GDP to unity. This serves as an identifying restriction in our estimation algorithm. Bai and Wang (2015) discuss minimal identifying assumptions for DFMs.

driftless random walks:

$$a_{j,t} = a_{j,t-1} + v_{a_{j,t}}, \qquad v_{a_{j,t}} \overset{iid}{\sim} N(0, \omega^2_{a,j}) \quad j = 1, \dots, r \tag{5}$$

$$\log \sigma_{\varepsilon_t} = \log \sigma_{\varepsilon_{t-1}} + v_{\varepsilon,t}, \qquad v_{\varepsilon,t} \overset{iid}{\sim} N(0, \omega^2_{\varepsilon}) \tag{6}$$

$$\log \sigma_{\eta_{i,t}} = \log \sigma_{\eta_{i,t-1}} + v_{\eta_{i,t}}, \qquad v_{\eta_{i,t}} \overset{iid}{\sim} N(0, \omega^2_{\eta,i}) \quad i = 1, \dots, n \tag{7}$$

where $a_{j,t}$ are the r time-varying elements in $\mathbf{a}_t$, and σ_{ε_t} and $\sigma_{\eta_{i,t}}$ capture the SV of the innovations to factor and idiosyncratic components. Our motivation for specifying the time-varying parameters as random walks is similar to Primiceri (2005). While in principle it is unrealistic model real GDP growth as a process that could wander in an unbounded way, as long as the variance of the process is small and the drift is considered to be operating for a finite period of time, the assumption is innocuous. Moreover, modeling a trend as a random walk is more robust to misspecification when the actual process is instead characterized by discrete breaks, whereas models with discrete breaks might not be robust to the true process being a random walk.[10] Finally, the random walk assumption also has the desirable feature that, unlike stationary models, confidence bands around forecasts of real GDP growth increase with the forecast horizon, reflecting uncertainty about the possibility of future breaks or drifts in long-run growth.

Note that a standard DFM is usually specified under two assumptions. First, the original data have been differenced appropriately so that both the factor and the idiosyncratic components can be assumed to be stationary. Second, it is assumed that the innovations in the idiosyncratic and common components are iid. In equations (1)-(7) we have relaxed these assumptions to allow for two novel features, a stochastic trend in the mean of selected series, and SV. By shutting down these features, we can recover the specifications previously proposed in the literature, which are nested in our framework. We obtain the DFM with SV

[10] We demonstrate this point with the use of Monte Carlo simulations, showing that a random walk trend in real GDP growth 'learns' quickly about a discrete break once it has occurred. On the other hand, the random walk does not detect a drift when there is not one, despite the presence of a large cyclical component. Online Appendix C provides a discussion and the full results of these simulations.

of Marcellino et al. (2014) if we shut down time-variation in the intercepts of the observables, i.e. set $r = m = 0$ and $\mathbf{c}_t = \mathbf{c}$. If we further shut down the SV, i.e. set $\omega^2_{a,j} = \omega^2_{\epsilon} = \omega^2_{\eta,i} = 0$, we obtain the specification of Banbura and Modugno (2014) and Banbura et al. (2012).

3.2 A Baseline Specification for Long-Run Growth

Equations (1) and (2) allow for stochastic trends in the mean of all or a subset of selected observables in $\mathbf{y}_t$. This paper focuses on tracking changes in the long-run growth rate of real GDP. For this purpose, the simplest specification of $\mathbf{c}_t$ is to include a time-varying intercept only in GDP and to set $\mathbf{B} = 1$. However, a number of empirical studies (e.g. Harvey and Stock, 1988, Cochrane, 1994, and Cogley, 2005) argue that incorporating information about consumption is informative about the permanent component in GDP as predicted by the permanent income hypothesis. The theory predicts that consumers, smoothing consumption throughout their lifetime, should react more strongly to permanent, as opposed to transitory, changes in income. As a consequence, looking at GDP and consumption data together will help separating growth into long-run and cyclical fluctuations.[11] Therefore, our baseline specification imposes that consumption and output grow at the same rate g_t in the long-run. On the contrary, we do not impose that investment also grows at this rate, as would be the case in the basic neoclassical growth model, since the presence of investment-specific technological change implies that real investment has a different low-frequency trend (Greenwood et al., 1997).

Formally, ordering real GDP and consumption growth first, and setting $m = 2$ and $r = 1$, this is represented as

$$\mathbf{a}_t = g_t, \quad \mathbf{B} = [1\ 1]' \tag{8}$$

Note that in this baseline specification we model time-variation only in the intercept for GDP and consumption while leaving it constant for the other observables. Of course it may

[11]While a strict interpretation of the permanent income hypothesis is rejected in the data, from an econometric point of view the statement applies as long as permanent changes are the main driver of consumption. See Cochrane (1994) for a very similar discussion.

be the case that some of the remaining $n - m$ series in $\mathbf{y}_t$ feature low frequency variation in their means. For instance, as mentioned above, this could be the case for investment. The key question is whether leaving it unspecified will affect the estimate of the long-run growth rate of GDP, which is our main object of interest. We ensure that this is not the case by allowing for persistence (and, in particular, we do not rule out unit roots) in the idiosyncratic components. If a series does feature a unit root which is not included in $\mathbf{a}_t$, its trend component will be absorbed by the idiosyncratic component. The choice of which elements to include in $\mathbf{a}_t$ therefore reflects the focus of a particular application.[12] Of course, if two series share the same underlying low-frequency component, and this is known with certainty, explicitly accounting for the shared low frequency variation will improve the precision of the estimation, but the risk of incorrectly including the trend is much larger than the risk of incorrectly excluding it. Therefore, in our baseline specification we include in $\mathbf{a}_t$ the intercept for GDP and consumption, while leaving any possible low-frequency variation in other series to be captured by the respective idiosyncratic components.[13]

An extension to include additional time-varying intercepts is straightforward through the flexible construction of $\mathbf{c}_t$ in equation (2). In fact, in Section 5 we explore how interest in the low-frequency movements of additional series leads to alternative choices for $\mathbf{a}_t$ and $\mathbf{B}$.[14]

[12] In principle, these unmodeled trends could still be recovered from our specification by applying a Beveridge-Nelson decomposition to its estimated idiosyncratic component. In practice, any low-frequency variation in the idiosyncratic component is likely to be obscured by a large amount of high frequency noise in the data and as result the extracted Beveridge-Nelson trend component will be imprecisely estimated, and as Morley et al. (2003) show, will not be smooth. In our specification, the elements of $\mathbf{a}_t$ are instead extracted directly, so that we are able to improve the extraction by imposing additional assumptions (e.g. smoothness) and prior beliefs (e.g. low variability) on its properties.

[13] We confirm this line of reasoning with a series of Monte Carlo experiments, in which data is generated from a system that features low-frequency movements in more series, which are left unmodeled in the estimation. Both in the case of series with independent trends and the case of series which share the trend of interest, the fact that they are left unmodeled has little impact on the estimate of the latter. Online Appendix C presents further discussion and the full results of these simulations.

[14] Note that the limiting case explicitly models time-varying intercept in all indicators, so that $m = r = n$ and $\mathbf{B} = \mathbf{I}_n$, i.e. an identity matrix of dimension n. See Creal et al. (2010) and Fleischman and Roberts (2011) for similar approaches. This setup would imply that the number of state variables increases with the number of observables, which severely increases the computational burden of the estimation, while offering little additional evidence with respect to the focus of this paper.

3.3 Dealing with Mixed Frequencies and Missing Data

Tracking activity in real time requires a model that can efficiently incorporate information from series measured at different frequencies. In particular, it must include both quarterly variables, such as the growth rate of real GDP, as well as more timely monthly indicators of real activity. Therefore, the model is specified at monthly frequency, and following Mariano and Murasawa (2003), the (observed) quarterly growth rates of a generic quarterly variable, x_t^q, can be related to the (unobserved) monthly growth rate x_t^m and its lags using a weighted mean. Specifically,

$$x_t^q = \frac{1}{3}x_t^m + \frac{2}{3}x_{t-1}^m + x_{t-2}^m + \frac{2}{3}x_{t-3}^m + \frac{1}{3}x_{t-4}^m, \tag{9}$$

and only every third observation of x_t^q is actually observed. Substituting the corresponding line of (1) into (9) yields a representation in which the quarterly variable depends on the factor and its lags. The presence of mixed frequencies is thus reduced to a problem of missing data in a monthly model.

Besides mixed frequencies, additional sources of missing data in the panel include: the "ragged edge" at the end of the sample, which stems from the non-synchronicity of data releases; missing data at the beginning of the sample, since some data series have been created or collected more recently than others; and missing observations due to outliers and data collection errors. Our Bayesian estimation method exploits the state space representation of the DFM and jointly estimates the latent factors, the parameters, and the missing data points using the Kalman filter (see Durbin and Koopman, 2012, for a textbook treatment).

3.4 State Space Representation and Estimation

The model features autocorrelated idiosyncratic components (see equation (4)). In order to cast it in state-space form, we include the idiosyncratic components of the quarterly variables in the state vector, and we redefine the system for the monthly indicators in terms of quasi-differences (see e.g. Kim and Nelson, 1999b, pp. 198-199, and Bai and Wang, 2015).[15] The model is estimated with Bayesian methods simulating the posterior distribution of parameters and factors using a Markov Chain Monte Carlo (MCMC) algorithm. We closely follow the Gibbs-sampling algorithm for DFMs proposed by Bai and Wang (2015), but extend it to include mixed frequencies, the time-varying intercept, and SV. The SVs are sampled using the approximation of Kim et al. (1998), which is considerably faster than the exact Metropolis-Hastings algorithm of Jacquier et al. (2002). Our complete sampling algorithm together with the details of the state space representation can be found in Section D of the Online Appendix.

4 Results for US Data

4.1 Data Selection

Our data set includes four key business cycle variables measured at quarterly frequency (output, consumption, investment and aggregate hours worked), as well as a set of 24 monthly indicators which are intended to provide additional information about cyclical developments in a timely manner.

The included quarterly variables are strongly procyclical and are considered key indicators of the business cycle (see e.g. Stock and Watson, 1999). Furthermore, theory predicts that they will be useful in disentangling low frequency movements from cyclical fluctuations in output growth. Indeed, as discussed in Section 3.2, the permanent income hypothesis predicts that consumption data will be particularly useful for the estimation of the long-run growth component, g_t.[16] On the other hand, investment and hours worked are very sensitive

[15]Since the quarterly variables are observed only every third month, we cannot take the quasi-difference for their idiosyncratic components, which are instead added as an additional state with the corresponding transition dynamics. Banbura and Modugno (2014) suggest including all of the idiosyncratic components as additional elements of the state vector. Our solution has the desirable feature that the number of state variables will increase with the number of quarterly variables, rather than the total number of variables, leading to a gain of computational efficiency.

[16]Due to the presence of faster technological change in the durable goods sector there is a downward trend in the relative price of durable goods. As a consequence, measured consumption grows faster than overall GDP. Following a long tradition in the literature (see e.g. Whelan, 2003), we construct a Fisher index of

to cyclical fluctuations, and thus will be particularly informative for the estimation of the common factor, f_t.[17]

The additional monthly indicators are crucial to our objective of disentangling in real time the cyclical and long-run components of GDP growth, since the quarterly variables are only available with substantial delay. In principle, a large number of candidate series are available to inform the estimate of f_t, and indirectly, of g_t. In practice, however, macroeconomic data series are typically clustered in a small number of broad categories (such as production, employment, or income) for which disaggregated series are available along various dimensions (such as economic sectors, demographic characteristics, or expenditure categories). The choice of which available series to include for estimation can therefore be broken into, first, a choice of which broad categories to include, and second, to which level and along which dimensions of disaggregation.

With regards to which broad categories of data to include, previous studies agree that prices, monetary and financial indicators are uninformative for the purpose of tracking real GDP, and argue for extracting a single common factor that captures real economic activity.[18] As for the possible inclusion of disaggregated series within each category, Boivin and Ng (2006) argue that the presence of strong correlation in the idiosyncratic components of disaggregated series of the same category will be a source of misspecification that can worsen the performance of the model in terms of in-sample fit and out-of-sample forecasting of key series.[19] Alvarez et al. (2012) investigate the trade-off between DFMs with very few

non-durables and services and use its growth rate as an observable variable in the panel. It can be verified that the ratio of consumption defined in this manner to real GDP displays no trend in the data, unlike the trend observed in the ratio of overall consumption to GDP.

[17]We define investment as a chain-linked aggregate of business fixed investment and consumption of durable goods, which is consistent with our treatment of consumption. In order to obtain a measure of hours for the total economy, we follow the methodology of Ohanian and Raffo (2012) and benchmark the quarterly series of hours in the non-farm business sector provided by the BLS to the annual estimates of hours in the total economy compiled by the Conference Boards Total Economy Database (TED). The TED series has the advantage of being comparable across countries (Ohanian and Raffo, 2012), which will be useful for our international results in Section 5.

[18]Giannone et al. (2005) conclude that that prices and monetary indicators do not contribute to the precision of GDP nowcasts. Banbura et al. (2012), Forni et al. (2003) and Stock and Watson (2003) find at best mixed results for financial variables.

[19]This problem is exacerbated by the fact that more detailed disaggregation levels and dimensions are

indicators, where the good large-sample properties of factor models are unlikely to hold, and those with a very large amount of indicators, where the problems above are likely to arise. They conclude that using a medium-sized panel with representative indicators of each category yields the best forecasting results.

The above considerations lead us to select 24 monthly indicators that include the high-level aggregates for all of the available broad categories that capture real activity, without overweighting any particular category. The complete list of variables contained in our data set is presented in Table 1. As the table shows, we include representative series of expenditure and income, the labor market, production and sales, foreign trade, housing and business and consumer confidence.[20] The inclusion of all the available monthly surveys is particularly important. Apart from being the most timely series available, these are unlikely to feature permanent shifts in their mean by construction, and have a high signal-to-noise ratio. They thus provide a clean signal to separate the cyclical component of GDP growth from its long-run counterpart. In Section 4.6 we explore sensitivity of our results to the size and composition of the data panel used.

Our panel spans the period January 1947 to March 2015. The start of our sample coincides with the year for which quarterly national accounts data are available from the Bureau of Economic Analysis. This enables us to study the evolution of long-run growth over the entire postwar period.[21]

available for certain categories of data, such as employment, meaning that the disaggregation will automatically 'tilt' the factor estimates towards that category.

[20]When there are multiple candidates for the high-level aggregate of a category, we include both. For example, we include employment as measured both by the establishment and household surveys, and consumer confidence as surveyed both by the Conference Board and the University of Michigan.

[21]We take full advantage of the Kalman filter's ability to deal with missing observations at any point in the sample, and we are able to incorporate series that become available substantially later than 1947, up to as late as 2007. Note that for consumption expenditures, monthly data became available in 1959, whereas quarterly data is available from 1947. In order to use all available data, we apply the polynomial in Equation (9) to the monthly data and treat the series as quarterly, with available observations for the last month of the quarter for 1947-1958 and for all months since 1959.

Table 1: Data series used in empirical analysis

	Type	Start Date	Transform.	Lag
Quarterly time series				
Real GDP	Expenditure & Inc.	Q2:1947	% QoQ Ann	26
Real Consumption (excl. durables)	Expenditure & Inc.	Q2:1947	% QoQ Ann	26
Real Investment (incl. durable cons.)	Expenditure & Inc.	Q2:1947	% QoQ Ann	26
Total Hours Worked	Labor Market	Q2:1948	% QoQ Ann	28
Monthly indicators				
Real Personal Income less Transfers	Expenditure & Inc.	Feb 59	% MoM	27
Industrial Production	Production & Sales	Jan 47	% MoM	15
New Orders of Capital Goods	Production & Sales	Mar 68	% MoM	25
Real Retail Sales & Food Services	Production & Sales	Feb 47	% MoM	15
Light Weight Vehicle Sales	Production & Sales	Feb 67	% MoM	1
Real Exports of Goods	Foreign Trade	Feb 68	% MoM	35
Real Imports of Goods	Foreign Trade	Feb 69	% MoM	35
Building Permits	Housing	Feb 60	% MoM	19
Housing Starts	Housing	Feb 59	% MoM	26
New Home Sales	Housing	Feb 63	% MoM	26
Payroll Empl. (Establishment Survey)	Labor Market	Jan 47	% MoM	5
Civilian Empl. (Household Survey)	Labor Market	Feb 48	% MoM	5
Unemployed	Labor Market	Feb 48	% MoM	5
Initial Claims for Unempl. Insurance	Labor Market	Feb 48	% MoM	4
Monthly indicators (soft)				
Markit Manufacturing PMI	Business Confidence	May 07	-	-7
ISM Manufacturing PMI	Business Confidence	Jan 48	-	1
ISM Non-manufacturing PMI	Business Confidence	Jul 97	-	3
NFIB Small Business Optimism Index	Business Confidence	Oct 75	Diff 12 M.	15
U. of Michigan: Consumer Sentiment	Consumer Confid.	May 60	Diff 12 M.	-15
Conf. Board: Consumer Confidence	Consumer Confid.	Feb 68	Diff 12 M.	-5
Empire State Manufacturing Survey	Business (Regional)	Jul 01	-	-15
Richmond Fed Mfg Survey	Business (Regional)	Nov 93	-	-5
Chicago PMI	Business (Regional)	Feb 67	-	0
Philadelphia Fed Business Outlook	Business (Regional)	May 68	-	0

Notes: % QoQ Ann refers to the quarter on quarter annualized growth rate, % MoM refers to $(y_t - y_{t-1})/y_{t-1}$ while Diff 12 M. refers to $y_t - y_{t-12}$. The last column shows the average publication lag, i.e. the number of days elapsed from the end of the period that the data point refers to until its publication by the statistical agency. All series were obtained from the Haver Analytics database.

4.2 Model Settings and Priors

The choice of the data set justifies the single-factor structure of the model. f_t can in this case be interpreted as a coincident indicator of real economic activity (see e.g. Stock and Watson, 1989, and Mariano and Murasawa, 2003). The number of lags in the polynomials $\phi(L)$ and $\rho(L)$ is set to $p = 2$ and $q = 2$ as in Stock and Watson (1989). We wish to impose as little prior information as possible, so we use uninformative priors for the factor loadings and the autoregressive coefficients of factors and idiosyncratic components. The variances of the innovations to the time-varying parameters, namely ω_a^2, ω_ε^2 and $\omega_{\eta,i}^2$ in equations (5)-(7) are however difficult to identify from the information contained in the likelihood alone. As the literature on Bayesian VARs documents, attempts to use non-informative priors for these parameters will in many cases produce posterior estimates which imply a relatively large amount of time-variation. While this will tend to improve the in-sample fit of the model it is also likely to worsen out-of-sample forecast performance. We therefore use priors to shrink these variances towards zero, i.e. towards the standard DFM which excludes time-varying long-run GDP growth and SV. In particular, for ω_a^2 we set an inverse gamma prior with one degree of freedom and scale equal to 0.001.[22] For ω_ϵ^2 and $\omega_{\eta,i}^2$ we set an inverse gamma prior with one degree of freedom and scale equal to 0.0001, closely following Cogley and Sargent (2005) and Primiceri (2005).[23] We estimate the model with 7000 replications of the Gibbs-sampling algorithm, of which the first 2000 are discarded as burn-in draws and the remaining ones are kept for inference.[24]

[22]To gain an intuition about this prior, note that over a period of ten years, this would imply that the random walk process of the long-run growth rate is expected to vary with a standard deviation of around 0.4 percentage points in annualized terms, which is a fairly conservative prior.

[23]We provide further explanations and address robustness to the choice of priors in Online Appendix F.

[24]Thanks to the efficient state space representation discussed above, the improvements in the simulation smoother proposed by Bai and Wang (2015), and other computational improvements we implemented, the estimation is very fast. Convergence is achieved after only 1500 iterations, which take less than 20 minutes in MATLAB using an Intel 3.6 GHz computer with 16GB of DDR3 Ram.

4.3 In-Sample Results

Panel (a) of Figure 2 plots the posterior median, together with the 68% and 90% posterior credible intervals of the long-run growth rate of real GDP. This estimate is conditional on the entire sample and accounts for both filtering and parameter uncertainty. Several features of our estimate of long-run growth are worth noting. While the growth rate is stable between 3% and 4% during the first decades of the postwar period, a slowdown is clearly visible from around the late 1960's through the 1970's, consistent with the "productivity slowdown" (Nordhaus, 2004). The acceleration of the late 1990's and early 2000's associated with the productivity boom in the IT sector (Oliner and Sichel, 2000) is also visible. Thus, until the middle of the decade of the 2000's, our estimate conforms well to the generally accepted narrative about fluctuations in potential growth.[25] More recently, after peaking at about 3.5% in 2000, the median estimate of the long-run growth rate has fallen to about 2% in early 2015, a more substantial decline than the one observed after the productivity slowdown of the 1970's. Moreover, the slowdown appears to have happened gradually since the start of the 2000's, with most of the decline having occurred before the Great Recession.[26] Interestingly, a small rebound is visible at the end of the sample, but long-run growth stands far below its postwar average of 3.2%, with the 90% posterior credible interval ranging from 1.5% to 2.5%.

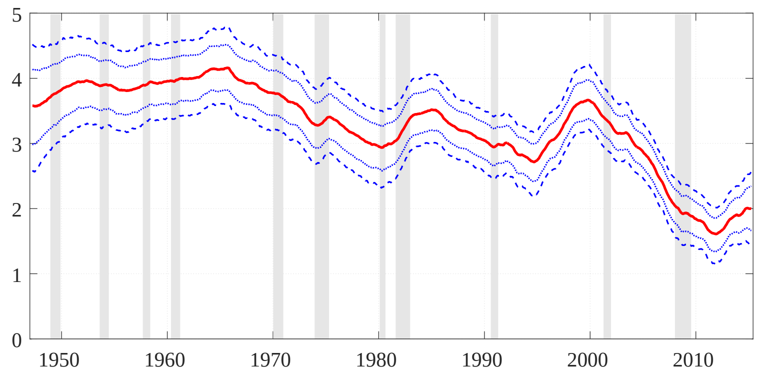
Panel (b) plots the time series of quarterly real GDP growth, together with the median posterior estimates of the common factor, aligned with the mean of real GDP growth. This plot highlights how the common factor captures the bulk of business-cycle frequency variation in output growth, while higher frequency, quarter-to-quarter variation is attributed to the

[25] Online Appendix G provides a comparison of our estimate with the Congressional Budget Office (CBO) measure of potential growth, with some additional discussion.

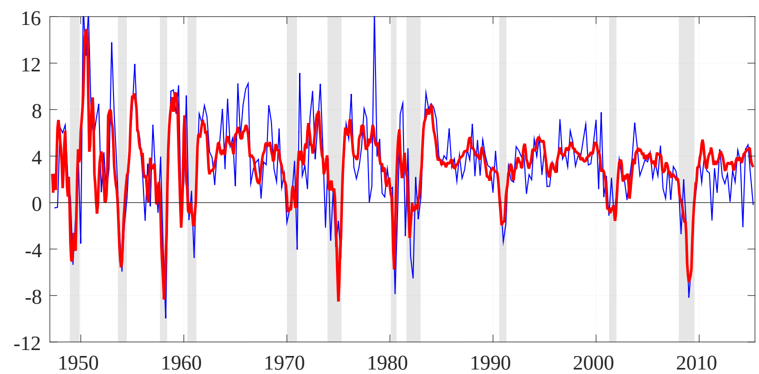
[26] In principle, it is possible that our choice of modeling long-run GDP growth as a random walk is hard-wiring into our results the conclusion that the decline happened in a gradual way. In experiments with simulated data, presented in Section C of the Online Appendix, we show that if changes in long-run growth occur in the form of discrete breaks rather than evolving gradually, the (one-sided) filtered estimates will exhibit a discrete jump at the moment of the break. Instead, for US data the filtered estimates of the long-run growth component also decline in a gradual manner (see Figure A.1 in Online Appendix A).

Figure 2: Trend, cycle and volatility: 1947-2015 (% Ann. Growth Rate)

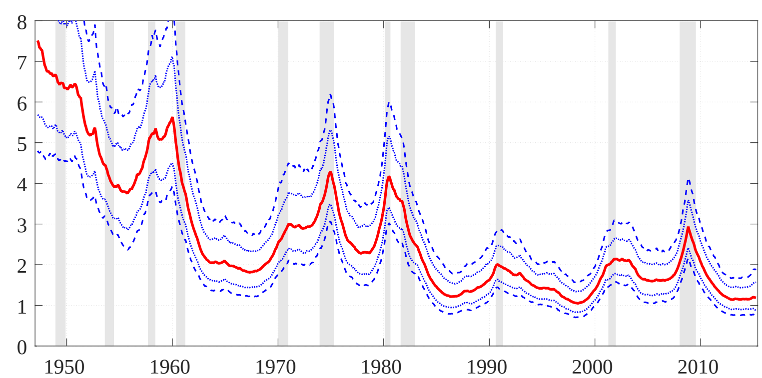
(a) Posterior estimate of long-run growth



(b) Posterior estimate of common factor vs. actual GDP growth



(c) Posterior estimate of common factor volatility



Note: Panel (a) displays the posterior median (solid), together with the 68% and 90% (dotted and dashed) posterior credible intervals of long-run real GDP growth. Panel (b) plots actual real GDP growth (thin) against the posterior median estimate of the common factor, aligned with the mean of real GDP growth (thick). Panel (c) presents the median (solid), the 68% and the 90% (dotted and dashed) posterior credible intervals of the volatility of the common factor, i.e the square root of $\text{var}(f_t) = \sigma^2_{\varepsilon,t}(1-\phi_2)/[(1+\phi_2)((1-\phi_2)^2-\phi_1^2)]$. Shaded areas represent NBER recessions.

idiosyncratic component. In the latter part of the sample, GDP growth is visibly below the factor, reflecting the decline in long-run growth.

The posterior estimate of the SV of the common factor is presented in Panel (c). It is clearly visible that volatility declines over the sample. The late 1940's and 1950's were extremely volatile, with a first large drop in volatility in the early 1960's. The Great Moderation is also clearly visible, with the average volatility pre-1985 being much larger than the average of the post-1985 sample. Notwithstanding the large increase in volatility during the Great Recession, our estimate of the common factor volatility since then remains consistent with the Great Moderation still being in place. This confirms the early evidence reported by Gadea-Rivas et al. (2014). It is clear from the figure that volatility spikes during recessions, a feature that brings our estimates close to the recent findings of Jurado et al. (2014) and Bloom (2014) relating to business-cycle uncertainty.[27] It appears that the random walk specification is flexible enough to capture cyclical changes in volatility as well as permanent phenomena such as the Great Moderation. Online Appendix A contains analogous charts for the volatilities of the idiosyncratic components of selected data series. Similar to the volatility of the common factor, many of the idiosyncratic volatilities present sharp increases during recessions.

The above results provide evidence that a significant decline in long-run US real GDP growth occurred over the last decade, and are consistent with a relatively gradual decline since the early 2000's. Our estimates show that the bulk of the slowdown from the elevated levels of growth at the turn of the century occurred before the Great Recession, which is consistent with the narrative of Fernald (2014) on the fading of the IT productivity boom. This recent decline is the largest movement in long-run growth observed in the postwar period.

[27]It is interesting to note that while in our model the innovations to the level of the common factor and its volatility are uncorrelated, the fact that increases in volatility are observed during recessions indicate the presence of negative correlation between the first and second moments, implying negative skewness in the distribution of the common factor. We believe a more explicit model of this feature is an important priority for future research.

4.4 Real-Time Results

As emphasized by Orphanides (2003), macroeconomic time series are heavily revised over time and in many cases these revisions contain valuable information that was not available at initial release. Therefore, it is important to assess, using the data available at each point in time, when the model detected the slowdown in long-run growth. For this purpose, we reconstruct our data set using vintages of data available from the Federal Reserve Bank of St. Louis ALFRED data base. Our aim is to replicate as closely as possible the situation of a decision-maker which would have applied our model in real time. We fix the start of our sample in 1947:Q1 and use an expanding out-of-sample window which starts on 11 January 2000 and ends on 30 June 2015. This is the longest possible window for which we are able to include the entire panel in Table 1 using fully real-time data. We then proceed by re-estimating the model each day in which new data are released.[28]

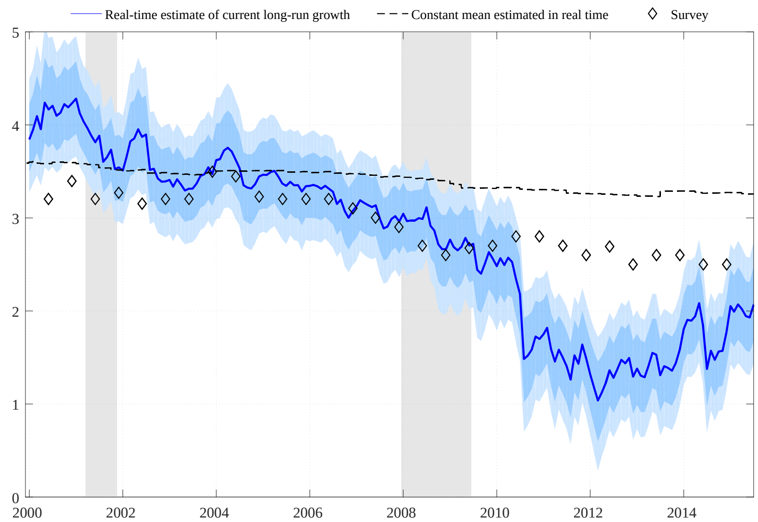
Figure 3 looks at the model's real-time assessment of long-run growth at various points in time. Panel (a) plots the real-time estimate of current long-run growth, with 68% and 90% uncertainty bands. For comparison, the panel also shows the median response to the Philadelphia Fed Livingston Survey of Professional Forecasters (SPF) on the average growth rate for the next 10 years, and the estimate of long-run growth from a model with a constant intercept for GDP growth. The latter estimate is also updated as new information arrives, but weighs all points of the sample equally. Panel (b) displays vintages of the median long-run growth estimate, using information available up to July of each year. The locus traced by the end point of each vintage corresponds to the current real-time estimate of Panel (a).

The evolution of the baseline model's estimate of long-run growth when estimated in real time declines gradually from a peak of about 4% in early 2000 to around 2.5% just after

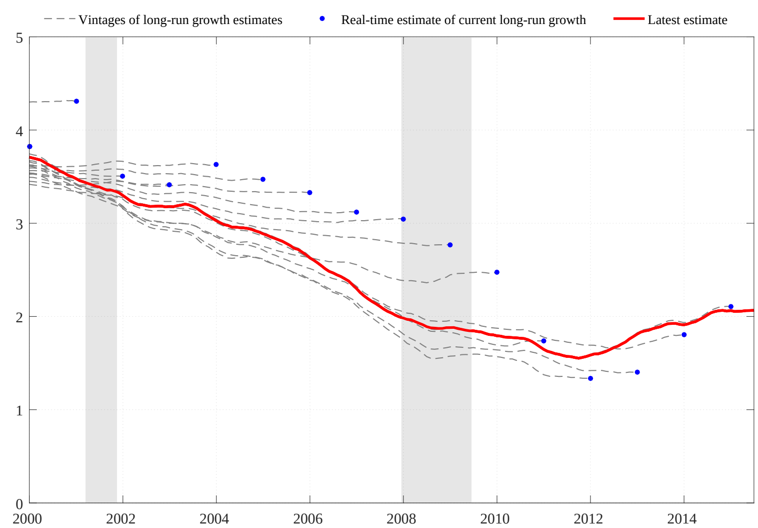
[28] In a few cases new indicators were developed after January 2000. For example, the Markit Manufacturing PMI survey is currently one of the most timely and widely followed indicators, but it started being conducted in 2007. In those cases, we append to the panel, in real time, the vintages of the new indicators as soon sufficient history is available. In the example of the PMI, this is the case since mid-2012. By implication, the number of indicators in our data panel grows when new indicators appear. Full details about the construction of the vintage database are available in Online Appendix E.

Figure 3: Long-Run GDP Growth Estimates in Real Time

(a) Evolution of the current assessment of long-run growth



(b) Selected vintages of long-run growth estimates using real-time data



Note: The figure presents results from re-estimating the model using the vintage of data available at each point in time from January 2000 to March 2015. The start of the estimation sample is fixed at Q1:1947. Panel (a) plots the median real-time estimate of current long-run growth over time. This is the locus traced by the end points of all vintages. The shaded areas around the solid line represent the 68th and 90th percentiles. The dashed line is the contemporaneous estimate of the historical average of real GDP growth. The diamonds are the median response to the Philadelphia Fed Livingston Survey of Professional Forecasters on the average growth rate for the next 10 years. Panel (b) displays the median estimate of long-run GDP growth for various vintages of data (dashed gray lines). The estimate of the latest vintage is shown as the solid thick line. Gray shaded areas represent NBER recessions in both panels.

the end of the Great Recession. From this time, the constant estimate shown in panel (a) is always outside of the 90% posterior bands. There is a sharp reassessment of long-run growth around July 2010, coinciding with the publication by the Bureau of Economic Analysis of the annual revisions to the National Accounts, which each year incorporate previously unavailable information for the previous three years. The revisions implied a substantial downgrade, in particular, to the growth of consumption in the first year of the recovery, from 2.5% to 1.6%, and instead allocated much of the growth in GDP during the recovery to inventory accumulation.[29] Reflecting the role of consumption as the most persistent and forward looking component of GDP, the estimate of long-run growth is downgraded sharply. Panel (b) shows how the 2010 revisions in fact trigger a re-interpretation of the years leading to the Great Recession. With the revised information, the bulk of the slowdown in long-run growth is now estimated to have occurred *before* the recession.[30] From 2010 onward, the model predicts a recovery that is extremely slow by historical standards. This is four years before the structural break test detected a statistically significant decline.[31] It is evident from the preceding discussion that revisions to past data by the BEA are an important source of changes to the long-run growth estimate in real time. Since the revision process is not modeled explicitly within the DFM, the in-sample results of Section 4.3 do not take into account the uncertainty stemming from future revisions. Interestingly, in the latest part of the sample, the estimate of long-run growth has recovered slightly to about 2% but this has been triggered by improvements in incoming data, rather than revisions to past vintages.

With regards to the SPF, it is noticeable that from 2003 to about 2010, the survey is remarkably similar to the model, but since then, the SPF forecast has continued to drift down only very slowly, standing at 2.5% as of mid-2015. It is noteworthy that, as pointed out by Stanley Fischer in the speech quoted in the introduction, during that period both

[29]See Online Appendix I for additional figures on the National Accounts revisions during this period.

[30]Indeed, the (one-sided) filtered estimate based on the latest vintage, which ignores the effect of data revisions, displays a more gradual pattern of decline (see Figure A.1 in Section A of the Online Appendix).

[31]A simpler specification that does not use consumption to inform the trend would detect the decline in long-run growth one year later, with additional revisions to past GDP in July 2011.

private and institutional forecasters systematically overestimated growth.

4.5 Implications for Nowcasting GDP

The standard DFM with constant long-run growth and constant volatility has been successfully applied to produce current quarter nowcasts of GDP (see Banbura et al., 2010, for a survey). Using our real-time US database, we carefully evaluate whether the introduction of time-varying long-run growth and SV into the DFM framework also improves the performance of the model along this dimension. We find that over the evaluation window 2000-2015 the model is at least as accurate at point forecasting, and significantly better at density forecasting than the benchmark DFM. We find that most of the improvement in density forecasting comes from correctly assessing the center and the right tail of the distribution, implying that the time-invariant DFM is assigning excessive probability to a strong recovery. In an evaluation sub-sample spanning the post-recession period, the relative performance of both point and density forecasts improves substantially, coinciding with the significant downward revision of the model's assessment of long-run growth. In fact, ignoring the variation in long-run GDP growth would have resulted in being on average around 1 percentage point too optimistic from 2009 to 2015.[32]

To sum up, the addition of the time-varying components not only provides a tool for decision-makers to update their knowledge about the state of long-run growth in real time. It also brings about a substantial improvement in short-run forecasting performance when the trend is shifting, without worsening the forecasts when the latter is relatively stable. The proposed model therefore provides a robust and timely methodology to track GDP when long-run growth is uncertain.

[32] Online Appendix H provides the full details of the forecast evaluation exercise.

4.6 Inspecting the Role of Data Set Size and Composition

In this paper we argue that the rich multivariate framework of a DFM will facilitate the extraction of the long-run growth component of GDP. The DFM will exploit the cross-sectional dimension, and not just the time series dimension in separating cycle from trend. It is interesting to quantify the advantage that the DFM provides over traditional trend-cycle decompositions, and to investigate the robustness of our main conclusions to alternative datasets of varying size and composition. In order to do so, we consider (1) a bivariate model with GDP and unemployment only (labeled "Okun"), (2) an intermediate model with GDP and the four additional variables often included in the construction of coincident indicators, see Mariano and Murasawa (2003) and Stock and Watson (1989) (labeled "MM03"), (3) our "Baseline" specification with 28 variables, and (4) an "Extended" model that uses disaggregated data for many of the headline series included in the baseline specification, totaling 155 variables.[33] Moreover, in order to investigate the gains associated with imposing additional structure to long-run GDP growth, for the last two specifications we also consider a version of the model that does not impose common long-run growth in GDP and consumption.

The top panel of Table 2 reports the mean point-estimates for each specification over selected subsamples.[34] In all cases, the results are consistent with a decline in the long-run growth rate in the last part of the sample. Quantitatively, most specifications are very close to the baseline, with the specifications that impose common long-run growth in GDP and consumption finding an earlier and sharper decline. The exception is the "Okun" specification which instead estimates a smaller increase in the mid 1990s as well as a larger

[33]As we argue in Section 4.1, the introduction of a large number of disaggregated series, even if related to real activity, is likely to lead to model misspecification whenever the sectoral data are not contemporaneously related. For the extended specification, we consider a solution to this problem which allows to maintain the parsimonious one factor structure. By extending the model to include lags of the factor in the observation equation, each variable can display heterogeneous responses to the common factor, and correlation between idiosyncratic components is reduced. Given that the extended model is heavily parameterized, we follow D'Agostino et al. (2015) in choosing priors that shrink the model towards the contemporaneous-only specification, which is nested in the extended case. Full details and the composition of the data set and the changes to the estimation in case of the extended model are provided in Online Appendix J.

[34]See Figure J.1 in Online Appendix J for a comparison of the results of each alternative specification with the baseline results over the entire sample.

decline in long-run growth in the past decade. It is noteworthy that the mean estimate of the extended specification is very close to that of the baseline.

Table 2:
COMPARISON OF RESULTS FOR ALTERNATIVE DATA SETS AND SPECIFICATIONS

			Baseline		Extended	
	Okun	*MM03*	*GDP only*	***GDP + C***	*GDP only*	*GDP + C*
Long-run growth						
1947-1972	3.9	3.5	3.6	**3.8**	3.6	3.9
1973-1995	3.2	3.4	3.1	**3.1**	3.2	3.2
1996-2007	2.6	3.2	3.1	**3.1**	3.0	3.1
2008-2015	1.6	2.5	2.4	**1.8**	2.2	1.7
End of Sample	1.3	2.4	2.3	**2.0**	2.1	2.0
Uncertainty: Long run						
Filtered	0.82	0.63	0.64	**0.56**	0.78	0.63
Smoothed	0.44	0.36	0.37	**0.35**	0.44	0.39
Uncertainty: Cycle						
Filtered	2.08	1.47	0.79	**0.76**	0.23	0.23
Smoothed	1.89	1.32	0.62	**0.60**	0.25	0.25

Notes: Each column presents the estimation results corresponding to the alternative models (data sets) considered in this section. The upper panel displays the posterior means of the long-run growth rate of real GDP, over selected subsamples. In the lower panel, the posterior uncertainty corresponding to both the long-run growth rate of real GDP, as well as the common factor are displayed. The uncertainty is calculated as an average over the entire sample.

The lower panel of Table 2 instead investigates the uncertainty around the mean estimates. The uncertainty around the long-run growth estimate declines as we move from the bivariate to the multivariate specifications, with most of the reduction happening once a handful of variables are included. On the other hand, when the panel is extended to include a large number of disaggregated series, the uncertainty increases.[35] While including a few

[35]We conjecture that as many more variables are added, the fit of the common factor to the cyclical component of GDP worsens. As a consequence, some cyclical variation of GDP spills over to the estimate of the long-run component. The uncertainty around the common factor, on the other hand, continues to decline.

key series, such as the ones in the specification of Mariano and Murasawa (2003) seems to already achieve the bulk of the reduction in uncertainty, it should be taken into account that those variables are available only with a relatively long publication lag, and subject to considerable revisions over time. Our proposed strategy of using an intermediate number of indicators, including the more timely and accurate surveys, is likely to lead to more satisfactory results in a real-time setting. Furthermore, the inclusion of the surveys is helpful in identifying the long-run growth rate, as those variables do not display a time-varying long-run mean by construction.

Overall this exercise highlights that the finding of a substantial decline in the long-run growth rate is confirmed across different specifications that use data sets of varying size and composition. The baseline specification, which uses an intermediate number of series including both hard data and surveys, leads to the lowest uncertainty around the long-run growth estimate, supporting the baseline choice of data set size and composition proposed in Section 4.1. Our results have important implications for trend-cycle decompositions of output, which usually include only a few cyclical indicators, generally inflation or variables that are direct inputs to the production function (see e.g. Gordon, 2014a or Reifschneider et al., 2013). As we show, greater precision of the trend component can be achieved by exploiting the common cyclical features of additional macroeconomic variables.[36]

5 Decomposing Movements in Long-Run Growth

In this section, we show how our model can be used to decompose the long-run growth rate of output into long-run movements in labor productivity and labor input. By doing this, we exploit the ability of the model to filter away cyclical variation and idiosyncratic noise and obtain clean estimates of underlying long-run trends. We see this exercise as a step towards giving an economically more meaningful interpretation to the movements in

[36]Basistha and Startz (2008) make a similar point, arguing that the inclusion of indicators that are informative about common cycles can help reduce the uncertainty around Kalman filter estimates of the long-run rate of unemployment (NAIRU).

long-run real GDP growth detected by our model.

GDP growth is by identity the sum of growth in output per hour and growth in total hours worked. It is therefore possible to split the long-run growth trend in our model into two orthogonal components such that this identity is satisfied in the long run. Here we make use of our flexible definition of $\mathbf{c}_t$ in equation (2). In particular, ordering the growth rates of real GDP, real consumption and total hours as the first three variables in $\mathbf{y}_t$, we define

$$\mathbf{a}_t = \begin{bmatrix} z_t \\ h_t \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}, \tag{10}$$

so that the model is specified with two time-varying components, the first of which loads output and consumption but not hours, and the second loads all three series. The first component is then by construction the long-run growth rate of labor productivity, while the second one captures low-frequency movements in labor input independent of productivity.[37] Given the relation in (10), the two components add up to the time-varying intercept in the baseline specification, i.e. $g_t = z_t + h_t$.[38] It follows from standard growth theory that our estimate of the long-run growth rate of labor productivity will capture both technological factors and other factors, such as capital deepening and labor quality.[39]

Figure 4 presents the results of the decomposition exercise for the US. Panel (a) plots the median posterior estimate of long-run real GDP growth and its labor productivity and total hours components. The posterior bands for long-run real GDP growth are included.

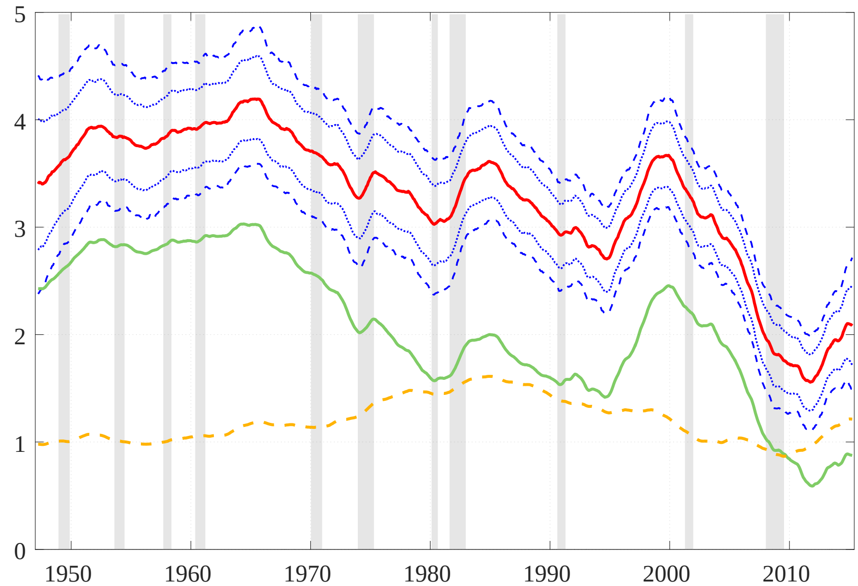
[37] z_t and h_t jointly follow random walks with diagonal covariance matrix as defined by equation (7). Restricting the covariance matrix is not necessary for estimation, but imposing it allows us to interpret the innovations to the trends as exogenous shocks to the long-run growth rates of the variables. The hours trend is therefore interpreted as those low-frequency movements in hours which are uncorrelated with labor productivity. Allowing for a full covariance matrix would yield trends that are linear combinations of the current ones, but would lack a clear economic interpretation.

[38] Since z_t and h_t are independent and add up to g_t, we set the prior on the scale of their variances to half of the one set in Section 4.2 on g_t. In addition, note that the cyclical movement in labor productivity is given by $(1 - \lambda_3)f_t$.

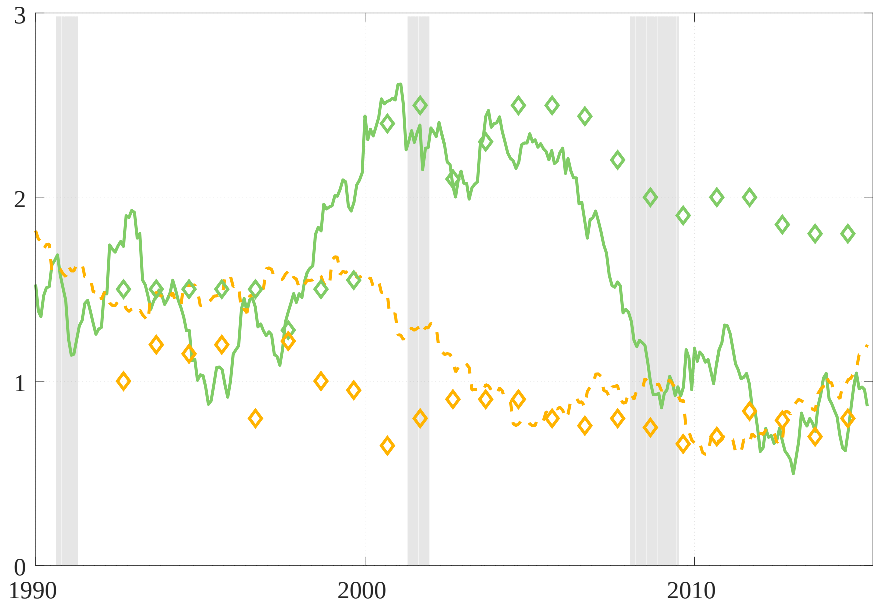
[39] Further decomposing z_t into technology and non-technology movements requires additional information to separately identify these components. One possibility, which we explore in Online Appendix K, is to use an independent measure of TFP to isolate technological factors. Note, however, that reliable data on capital input, labor quality, or estimates of TFP are not available in real time, making the focus on long-run labor productivity more appealing in a real-time setting.

Figure 4: Decomposition of Long-run US Output Growth

(a) Posterior median estimates of decomposition



(b) Filtered estimates of long-run growth components



Note: Panel (a) plots the posterior median (solid black), together with the 68% and 90% (doted and dashed gray) posterior credible intervals of long-run GDP growth and the posterior median of both long-run labor productivity growth and long-run total hours growth (crossed markers and circled markers). Panel (b) plots the filtered estimates of these two components, i.e. $\hat{z}_{t|t}$ and $\hat{h}_{t|t}$, since 1990. For comparison, the corresponding forecasts from the SPF are plotted (diamonds and squares). The SPF forecast for total hours is obtained as the difference between the forecasts for real GDP and labor productivity.

The time series evolution conforms very closely to the narrative of Fernald (2014), with a pronounced boom in labor productivity in the mid-1990's and a subsequent fall in the 2000's clearly visible. The decline in the 2000's is relatively sudden while the 1970's slowdown appears as a more gradual phenomenon starting in the late 1960's. Furthermore, the results reveal that during the 1970's and 1980's the impact of the productivity slowdown on output growth was partly masked by a secular increase in hours, probably reflecting increases in the working-age population as well as labor force participation (see e.g. Goldin, 2006). Focusing on the period since 2000, labor productivity accounts for almost the entire decline.[40] This contrasts explanations by which slow labor force growth has been a drag on GDP growth. When taking away the cyclical component of hours and focusing solely on its long-run component, the contribution of hours has, if anything, accelerated since the Great Recession. Panel (b) presents the filtered estimates of the two components, i.e. the output of the Kalman Filter which uses data only up to each point in time. For comparison, the corresponding SPF forecasts are included. Most notably, this plot reveals that starting around 2005 a relatively sharp revision to labor productivity drives the decline in long-run output growth.[41] Interestingly, the professional forecasters have been very slow in incorporating the productivity slowdown into their long-run forecasts. This delay explains their persistent overestimation of GDP growth since the recession.

It is interesting to compare the results of our decomposition exercise to similar approaches in the literature, in particular Gordon (2010, 2014a) and Reifschneider et al. (2013). Like us, they specify a state space model with a common cyclical component and use the 'output identity' to decompose the long-run growth rate of GDP into underlying drivers. A key difference resides in the Bayesian estimation of the model, which enables us to impose a conservative prior on the variance of the long-run growth component that helps avoiding

[40] In Online Appendix K we extend the analysis to decompose the labor productivity trend into long-run TFP and non-technological forces. We find that TFP accounts for virtually all of the slowdown.

[41] In an additional figure, provided in Section A of the Online Appendix, we plot 5,000 draws from the joint posterior distribution of the variances of the innovations to the labor productivity and hours components. This analysis confirms the conclusion from the discussion here that changes in labor productivity, rather than in labor input, are the key driver of low frequency movements in real GDP growth.

over-fitting the data. Furthermore, the inclusion of SV in the cyclical component helps to prevent unusually large cyclical movements from contaminating the long-run estimate. Another important difference is that we use a larger amount of information, including key cyclical indicators like industrial production, sales, and business surveys, which are generally not included in a production function approach. This allows us to retrieve a timely and precise estimate of the cyclical component and, as a consequence, to reduce the uncertainty that is inherent to any trend-cycle decomposition of the data, as discussed in Section 4.6. As a result, we obtain a substantially less pessimistic estimate of the long-run growth of GDP than these studies in the latest part of the sample. For instance, Gordon (2014a) reports a long-run GDP growth estimate below 1% for the end of the sample, whereas our median estimate stands at around 2%.[42]

5.1 International Evidence

To gain an international perspective on our results, we estimate the DFM for the other G7 economies and perform the decomposition exercise for each of them.[43] The median posterior estimates of the labor productivity and labor input trends are displayed in Figure 5. Labor productivity, displayed in Panel (a), plays again the key role in determining movements in long-run growth. In the Western European economies and Japan, the elevated growth rates of labor productivity prior to the 1970's reflect the rebuilding of the capital stock from the destruction from World War II, and ended as these economies converged towards US levels of output per capita. The labor productivity profile of Canada broadly follows that of the US, with a slowdown in the 1970's and a temporary mild boom during the late 1990's.

[42] The results for a bivariate model of GDP and unemployment, which we have discussed in Section 4.6 show that the current long-run growth estimate is 1.3%, close to Gordon (2014a).

[43] Details on the specific data series used for each country are available in Online Appendix E. For hours, we again follow the methodology of Ohanian and Raffo (2012). In the particular case of the UK, the quarterly series for hours displays a drastic change in its stochastic properties in the early 1990's owing to a methodological change in the construction by the ONS, as confirmed by the ONS LFS manual. We address this issue by using directly the annual series from the TED, which requires an appropriate extension of equation (9) to annual variables (see Banbura et al. 2012). To avoid weak identification of h_t for the UK, we truncate our prior on its variance to discard values which are larger than twice the maximum posterior draw of the case of the other countries.

Interestingly, this acceleration in the 1990's did not occur in Western Europe and Japan.[44] The UK displays a decline in labor productivity similar to the US. This "productivity puzzle" has been debated extensively in the UK (see e.g. Pessoa and Van Reenen, 2014). It is interesting to note that the two countries which experienced a more severe financial crisis, the US and the UK, appear to be the ones with greatest declines in productivity since the early 2000's, similar to the evidence documented in Reinhart and Rogoff (2009).

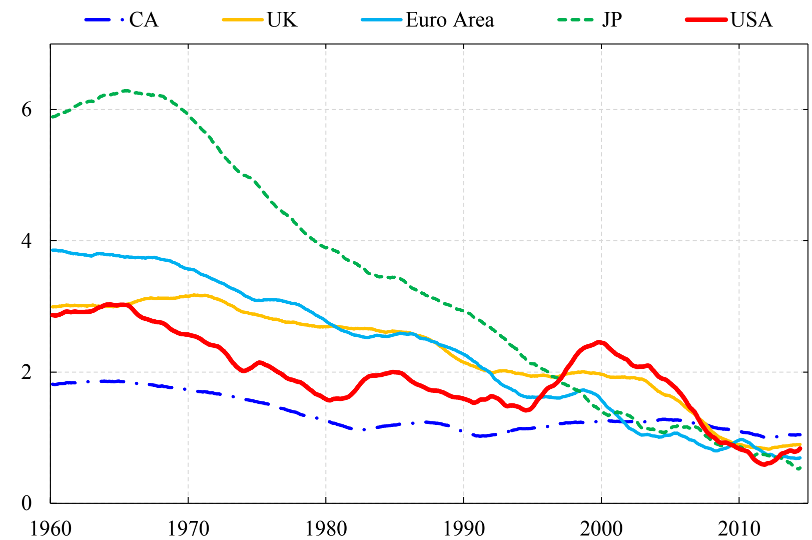
Panel (b) displays the movements in long-run hours worked identified by equation (10). The contribution of this component to overall long-run output growth varies considerably across countries. However, within each country it is more stable over time than the productivity component, which is in line with our findings for the US. Indeed, the extracted long-run trend in total hours includes various potentially offsetting forces that can lead to changes in long-run output growth. In any case, the results of our decomposition exercise indicate that after using the DFM to remove business-cycle variation in hours and output, the decline in long-run GDP growth that has been observed in the advanced economies since the early 2000's is entirely accounted for by a decline in the labor productivity trend. Finally, it is interesting to note that for the countries in the sample long-run productivity growth appears to converge in the cross section, while there is no evidence of convergence in the long-run growth of hours.[45]

[44]On the lost decade in Japan, see Hayashi and Prescott (2002). Gordon (2004) examines the absence of the IT boom in Europe.

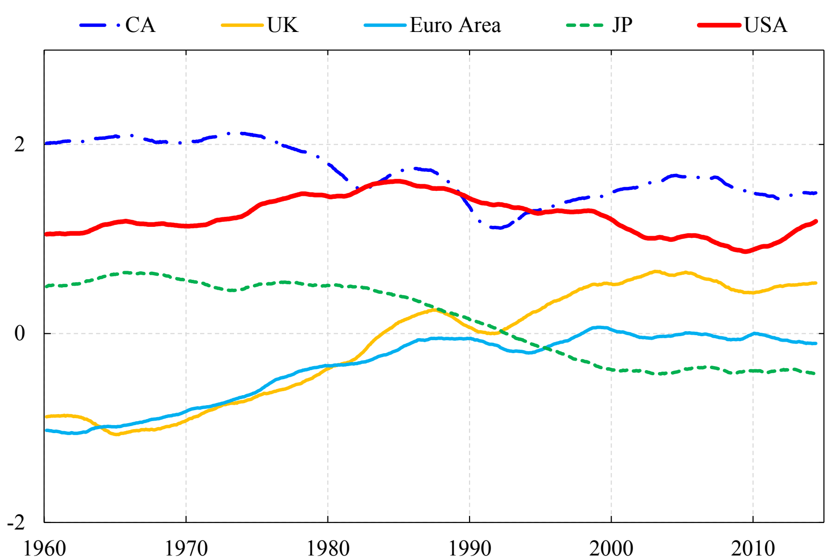
[45]Similar evidence for emerging economies has been recently presented by Pritchett and Summers (2014). Their evidence refers to convergence of overall GDP growth rates, whereas ours indicates that convergence in productivity growth appears to be the dominant source of convergence.

Figure 5: Decomposition for Other Advanced Economies

(a) Long-run Labor Productivity



(b) Long-run Labor Input



Note: Panel (a) displays the posterior median of long-run labor productivity across advanced economies. Panel (b) plots the corresponding estimates of long-run total hours worked. In both panels, 'Euro Area' represents a weighted average of Germany, Italy and France.

6 Concluding Remarks

The sluggish recovery from the Great Recession has raised the question whether the long-run growth rate of US real GDP is now lower than it has been on average over the postwar period. We have presented a dynamic factor model that allows for both changes in long-run GDP growth and stochastic volatility. Estimating the model with Bayesian methods, we provide evidence that long-run growth of US GDP displays a gradual decline after the turn of the century, moving from its peak of 3.5% to about 2% in 2015. Using real-time vintages of data we demonstrate the model's ability to track GDP in a timely and reliable manner. By the summer of 2010 the model would have concluded that a significant decline in long-run growth was behind the slow recovery, therefore substantially improving the real-time tracking of GDP by explicitly taking into account the uncertainty surrounding long-run growth. Finally, we discuss the drivers of movements in long-run output growth through the lens of our model by decomposing it into the long-run growth rates of labor productivity and labor input. Using data for both the US and other advanced economies our model points to a global slowdown in labor productivity as the main driver of weak growth in recent years, extending the narrative of Fernald (2014) to other economies. Studying the deep causes of the secular decline in growth is an important priority for future research.

References

Alvarez, R., Camacho, M., and Perez-Quiros, G. (2012). Finite sample performance of small versus large scale dynamic factor models. CEPR Discussion Papers 8867, C.E.P.R. Discussion Papers.

Auerbach, A. J. (2011). Long-term fiscal sustainability in major economies. BIS Working Papers 361, Bank for International Settlements.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–68.

Bai, J. and Wang, P. (2015). Identification and Bayesian Estimation of Dynamic Factor Models. *Journal of Business & Economic Statistics*, 33(2):221–240.

Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2012). Now-casting and the real-time data flow. Working Papers ECARES 2012-026, ULB.

Banbura, M., Giannone, D., and Reichlin, L. (2010). Nowcasting. CEPR Discussion Papers 7883, C.E.P.R. Discussion Papers.

Banbura, M. and Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160.

Basistha, A. and Startz, R. (2008). Measuring the NAIRU with Reduced Uncertainty: A Multiple-Indicator Common-Cycle Approach. *The Review of Economics and Statistics*, 90(4):805–811.

Benati, L. (2007). Drift and breaks in labor productivity. *Journal of Economic Dynamics and Control*, 31(8):2847–2877.

Beveridge, S. and Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics*, 7(2):151–174.

Bloom, N. (2014). Fluctuations in Uncertainty. *Journal of Economic Perspectives*, 28(2):153–76.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Clark, P. K. (1987). The Cyclical Component of U.S. Economic Activity. *The Quarterly Journal of Economics*, 102(4):797–814.

Cochrane, J. H. (1994). Permanent and Transitory Components of GNP and Stock Prices. *The Quarterly Journal of Economics*, 109(1):241–65.

Cogley, T. (2005). How fast can the new economy grow? A Bayesian analysis of the evolution of trend growth. *Journal of Macroeconomics*, 27(2):179–207.

Cogley, T. and Sargent, T. J. (2005). Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S. *Review of Economic Dynamics*, 8(2):262–302.

Creal, D., Koopman, S. J., and Zivot, E. (2010). Extracting a robust US business cycle using a time-varying multivariate model-based bandpass filter. *Journal of Applied Econometrics*, 25(4):695–719.

D'Agostino, A., Giannone, D., Lenza, M., and Modugno, M. (2015). Nowcasting business cycles: A bayesian approach to dynamic heterogeneous factor models. Technical report.

Del Negro, M. and Otrok, C. (2008). Dynamic factor models with time-varying parameters:measuring changes in international business cycles. NYFed Staff Rep. 326.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods: Second Edition*. Oxford University Press.

Edge, R. M., Laubach, T., and Williams, J. C. (2007). Learning and shifts in long-run productivity growth. *Journal of Monetary Economics*, 54(8):2421–2438.

Evans, M. D. D. (2005). Where Are We Now? Real-Time Estimates of the Macroeconomy. *International Journal of Central Banking*, 1(2).

Fernald, J. (2014). Productivity and potential output before, during, and after the great recession. *NBER Macroeconomics Annual 2014*, 29.

Fernald, J. G. and Jones, C. I. (2014). The Future of US Economic Growth. *American Economic Review*, 104(5):44–49.

Fleischman, C. A. and Roberts, J. M. (2011). From many series, one cycle: improved estimates of the business cycle from a multivariate unobserved components model. Finance and Economics Discussion Series 2011-46, Fed Board of Governors.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening The Black Box: Structural Factor Models With Large Cross Sections. *Econometric Theory*, 25(05):1319–1347.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2003). Do financial variables help forecasting inflation and real activity in the euro area? *Journal of Monetary Economics*, 50(6):1243–1255.

Gadea-Rivas, M. D., Gomez-Loscos, A., and Perez-Quiros, G. (2014). The Two Greatest. Great Recession vs. Great Moderation. WP Series 1423, Bank of Spain.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In *Latent Variables in Socio-Economic Models*. North-Holland.

Giannone, D., Reichlin, L., and Sala, L. (2006). Vars, common factors and the empirical validation of equilibrium business cycle models. *Journal of Econometrics*, 132(1):257–279.

Giannone, D., Reichlin, L., and Small, D. (2005). Nowcasting gdp and inflation: the real-time informational content of macroeconomic data releases. Technical report.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. *AEA Papers and Proceedings*, May 2006.

Gordon, R. J. (2004). Why was europe left at the station when america's productivity locomotive departed? Working Paper 10661, National Bureau of Economic Research.

Gordon, R. J. (2010). Okun's Law and Productivity Innovations. *American Economic Review*, 100(2):11–15.

Gordon, R. J. (2014a). A new method of estimating potential real gdp growth: Implications for the labor market and the debt/gdp ratio. NBER Working Paper 20423.

Gordon, R. J. (2014b). The Demise of U.S. Economic Growth: Restatement, Rebuttal, and Reflections. NBER Working Papers 19895.

Greenwood, J., Hercowitz, Z., and Krusell, P. (1997). Long-Run Implications of Investment-Specific Technological Change. *American Economic Review*, 87(3):342–62.

Hansen, B. E. (1992). Testing for parameter instability in linear models. *Journal of Policy Modeling*, 14(4):517–533.

Harvey, A. C. (1985). Trends and Cycles in Macroeconomic Time Series. *Journal of Business & Economic Statistics*, 3(3):216–27.

Harvey, A. C. and Stock, J. H. (1988). Continuous time autoregressive models with common stochastic trends. *Journal of Economic Dynamics and Control*, 12(2-3):365–384.

Hayashi, F. and Prescott, E. C. (2002). The 1990s in japan: A lost decade. *Review of Economic Dynamics*, 5(1):206 – 235.

Hodrick, R. J. and Prescott, E. C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, 29(1):1–16.

Jacquier, E., Polson, N. G., and Rossi, P. E. (2002). Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics*, 20(1):69–87.

Jorgenson, D. W., Ho, M. S., and Stiroh, K. J. (2006). Potential Growth of the U.S. Economy: Will the Productivity Resurgence Continue? *Business Economics*, 41(1):7–16.

Jurado, K., Ludvigson, S. C., and Ng, S. (2014). Measuring uncertainty. *American Economic Review, Forthcoming.*

Kim, C.-J. and Nelson, C. R. (1999a). Has The U.S. Economy Become More Stable? A Bayesian Approach Based On A Markov-Switching Model Of The Business Cycle. *The Review of Economics and Statistics*, 81(4):608–616.

Kim, C.-J. and Nelson, C. R. (1999b). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications.* The MIT Press.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65(3):361–93.

Luciani, M. and Ricci, L. (2014). Nowcasting Norway. *International Journal of Central Banking, Forthcoming.*

Luo, S. and Startz, R. (2014). Is it one break or on going permanent shocks that explains u.s. real gdp? *Journal of Monetary Economics*, 66:155–163.

Marcellino, M., Porqueddu, M., and Venditti, F. (2014). Short-term gdp forecasting with a mixed frequency dynamic factor model with stochastic volatility. *Journal of Business & Economic Statistics, Forthcoming.*

Mariano, R. S. and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18(4):427–443.

McConnell, M. M. and Perez-Quiros, G. (2000). Output fluctuations in the united states: What has changed since the early 1980's? *American Economic Review*, 90(5):1464–1476.

Morley, J. C., Nelson, C., and Zivot, E. (2003). Why are the beveridge-nelson and unobserved-components decompositions of gdp so different? *The Review of Economics and Statistics*, 85(2):235–243.

Nelson, C. R. and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, 10(2):139–162.

Nordhaus, W. (2004). Retrospective on the 1970s Productivity Slowdown. Nber working papers, National Bureau of Economic Research, Inc.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84(405):223–230.

Ohanian, L. E. and Raffo, A. (2012). Aggregate hours worked in OECD countries: New measurement and implications for business cycles. *Journal of Monetary Economics*, 59(1):40–56.

Oliner, S. D. and Sichel, D. E. (2000). The Resurgence of Growth in the Late 1990s: Is Information Technology the Story? *Journal of Economic Perspectives*, 14(4):3–22.

Orphanides, A. (2003). The quest for prosperity without inflation. *Journal of Monetary Economics*, 50(3):633–663.

Pain, N., Lewis, C., Dang, T.-T., Jin, Y., and Richardson, P. (2014). OECD Forecasts During and After the Financial Crisis: A Post Mortem. OECD Economics Department Working Papers 1107, OECD Publishing.

Perron, P. and Wada, T. (2009). Let's take a break: trends and cycles in us real gdp. *Journal of Monetary Economics*, 56:749–765.

Pessoa, J. P. and Van Reenen, J. (2014). The uk productivity and jobs puzzle: Does the answer lie in wage flexibility? *The Economic Journal*, 124(576):433–452.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852.

Pritchett, L. and Summers, L. H. (2014). Asiaphoria Meets Regression to the Mean. NBER Working Papers 20573, National Bureau of Economic Research, Inc.

Reifschneider, D., Wascher, W., and Wilcox, D. (2013). Aggregate supply in the United States: recent developments and implications for the conduct of monetary policy. Finance and Economics Discussion Series 2013-77, Board of Governors of the Fed.

Reinhart, C. M. and Rogoff, K. S. (2009). The Aftermath of Financial Crises. *American Economic Review*, 99(2):466–72.

Sargent, T. J. and Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. Working Papers 55, Federal Reserve Bank of Minneapolis.

Sims, C. A. (2012). Comment to stock and watson (2012). *Brookings Papers on Economic Activity*, Spring:81–156.

Stock, J. and Watson, M. (2012). Disentangling the channels of the 2007-09 recession. *Brookings Papers on Economic Activity*, Spring:81–156.

Stock, J. H. and Watson, M. W. (1989). New Indexes of Coincident and Leading Economic Indicators. In *NBER Macroeconomics Annual 1989, Volume 4*, NBER Chapters, pages 351–409. National Bureau of Economic Research, Inc.

Stock, J. H. and Watson, M. W. (1999). Business cycle fluctuations in us macroeconomic time series. In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics*, chapter 1, pages 3–64. Elsevier.

Stock, J. H. and Watson, M. W. (2002). Forecasting Using Principal Components From a

Large Number of Predictors. *Journal of the American Statistical Association*, 97:1167–1179.

Stock, J. H. and Watson, M. W. (2003). Has the business cycle changed and why? *NBER Macroeconomics Annual 2002*, 17:159–230.

Summers, L. (2014). Secular stagnation. IMF Economic Forum: Policy Responses to Crises, speech delivered at the IMF Annual Research Conference, November 8th.

Whelan, K. (2003). A Two-Sector Approach to Modeling U.S. NIPA Data. *Journal of Money, Credit and Banking*, 35(4):627–56.

Online Appendix to
"Tracking the Slowdown in Long-Run GDP Growth"

by Juan Antolin-Diaz, Thomas Drechsel and Ivan Petrella

Contents

A Additional Figures **3**

B Full Results of Structural Break Tests **6**

B.1 Nyblom Test 6

B.2 Bai and Perron Test 7

C Monte Carlo Evidence **8**

C.1 Setup for Monte Carlo simulations 8

C.2 Results: Sensitivity of random walk specification 9

C.3 Results: Sensitivity to confounding time-variation 14

D Details on Estimation Procedure **19**

D.1 Construction of the State Space System 19

D.2 Details of the Gibbs Sampler 21

D.3 Implementing linear restrictions on the factor loadings 22

E Details on the Construction of the Data Base **24**

E.1 US (Vintage) Data Base 24

E.2 Data Base for Other G7 Economies 27

F Choice of Priors **33**

F.1 Robustness checks on prior choice 33

G Comparison with CBO Measure of Potential Output **37**

H Details About the Forecast Evaluation 38

H.1 Setup . . . 38

H.2 Point Forecast Evaluation . . . 39

H.3 Density Forecast Evaluation . . . 39

I Case Study - The Decline of The Long-Run Growth Estimate in Mid-2010 and Mid-2011 45

J Inspecting Data Set Size and Composition - More Details 47

J.1 Extended Model: Estimation Using a Very Large Panel . . . 47

J.1.1 Construction of the Extended Data Set . . . 47

J.1.2 Extended Model Specification . . . 47

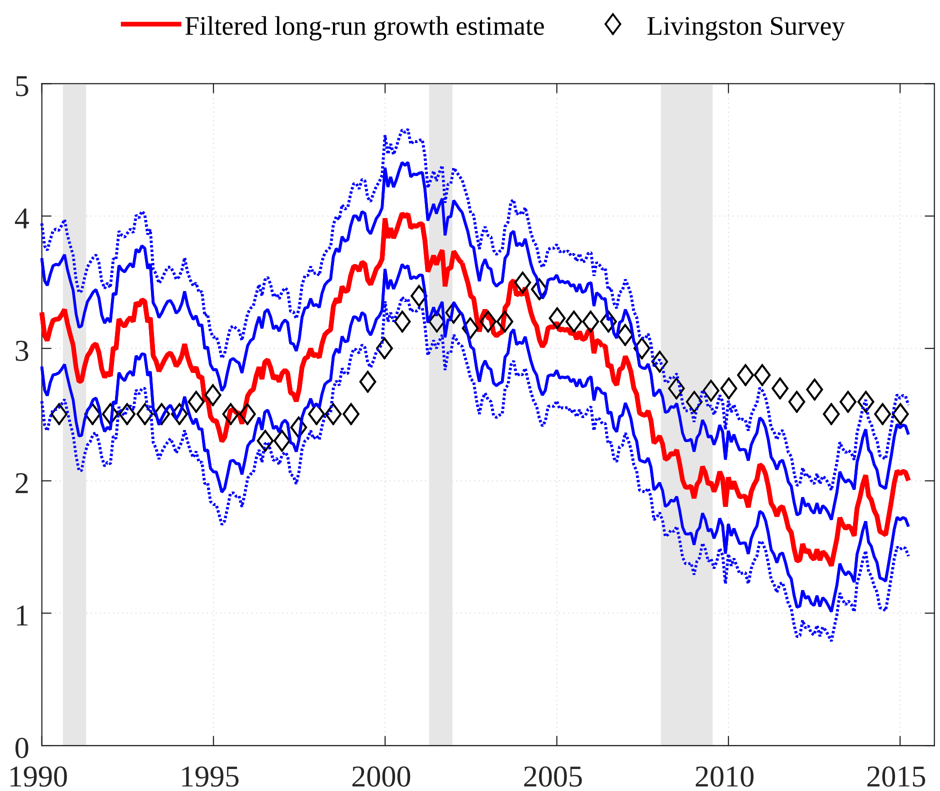
J.1.3 Priors and Model Settings . . . 48

J.2 Results Across Alternative Data Sets . . . 49

K A Growth Accounting Exercise 51

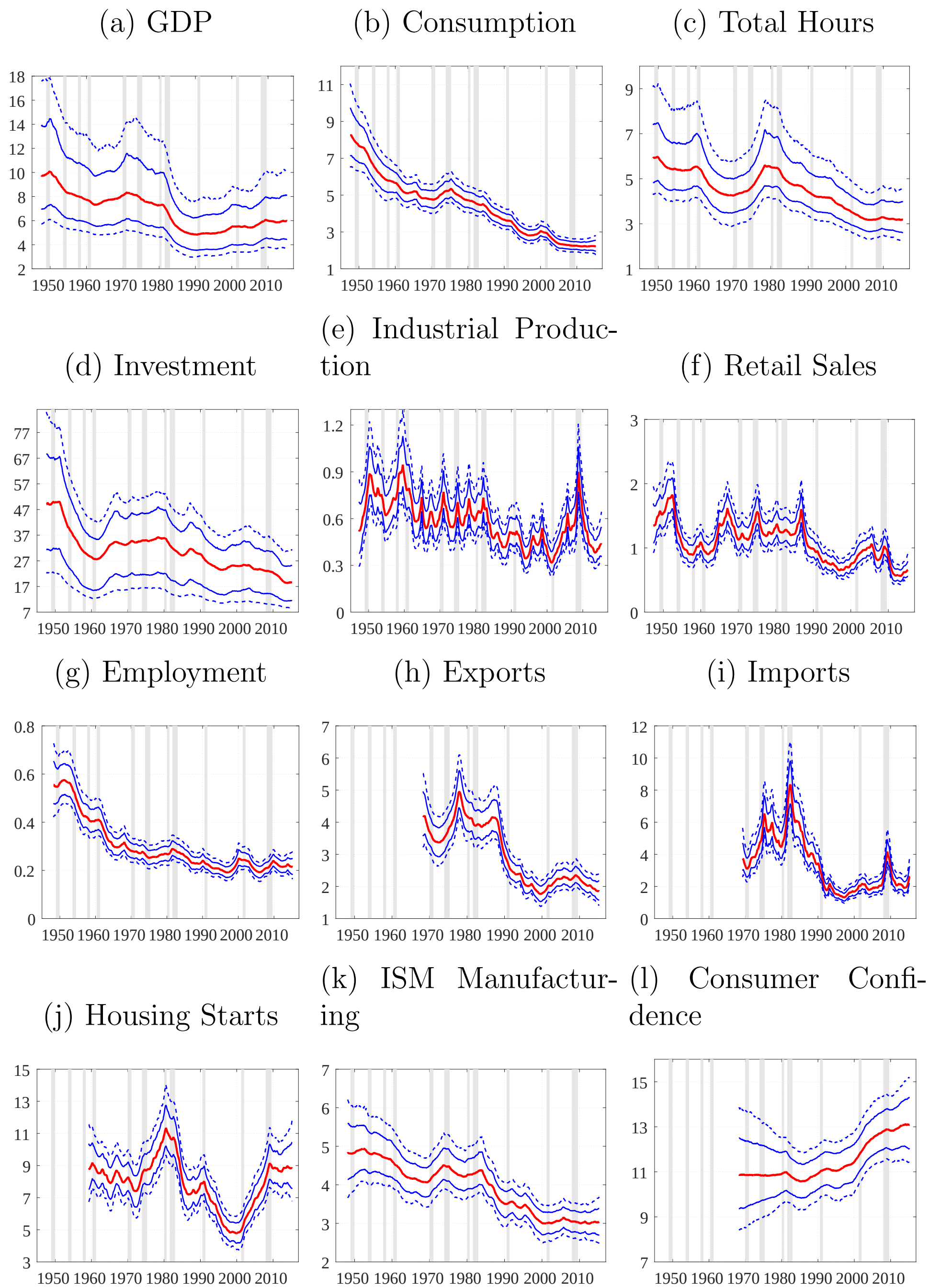
A Additional Figures

Figure A.1: Filtered estimate of long-run growth



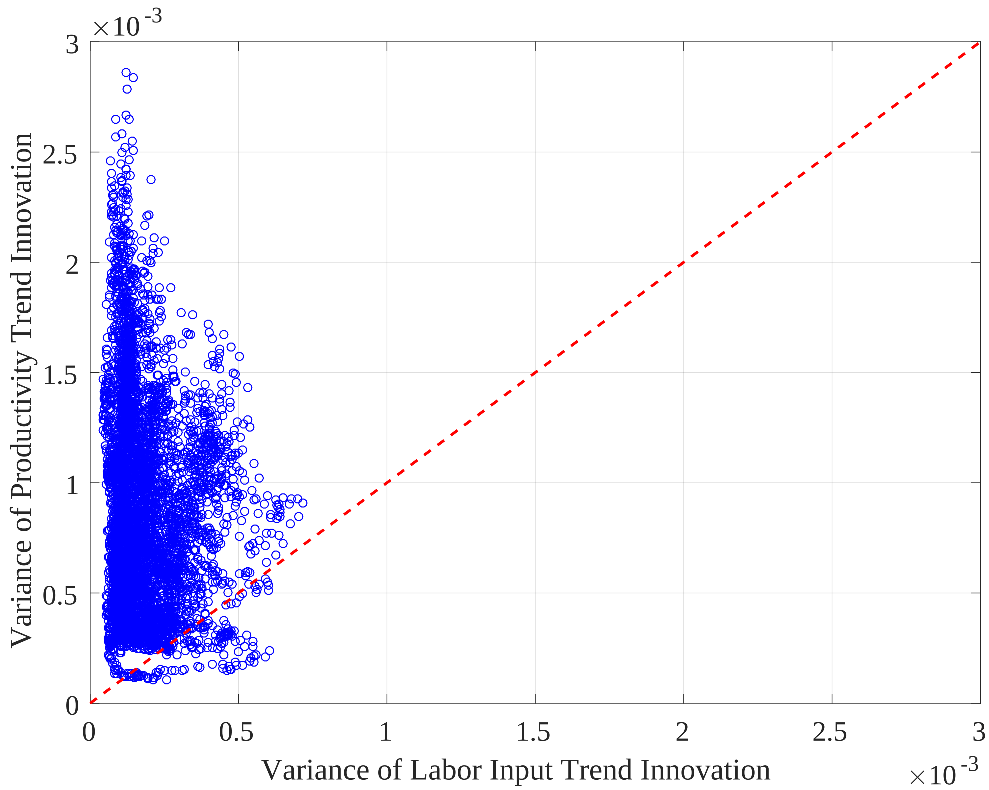
Note: The solid red line is the filtered estimate of the long-run GDP growth rate, $\hat{g}_{t|t}$, using the vintage of National Accounts available as of March 2015. The solid and dotted blue lines capture the corresponding 68% and 90% posterior bands. The black diamonds represent the real-time mean forecast from the Livingston Survey of Professional Forecasters of the average GDP growth rate for the subsequent 10 years.

Figure A.2: Stochastic Volatility of Selected Idiosyncratic Components



Note: Each panel presents the median (solid red), the 68% and the 90% (solid and dashed blue) posterior credible intervals of the volatility of the idiosyncratic component of selected variables. Shaded areas represent NBER recessions. Similar charts for other variables are available upon request.

Figure A.3: Joint Posterior Distribution of Growth Component Innovation Variances



Note: The figure plots 5,000 draws of the joint posterior distribution of the variances of innovations to the labor productivity and hours component. The dashed red line is the 45°line. Under the equal-variance prior the draws would be equally distributed above and below this line. The fact that the bulk of draws lie above indicates that changes in long-run labor productivity drive the variation in long-run output.

B Full Results of Structural Break Tests

B.1 Nyblom Test

Table B.1 reports the result for the Nyblom (1989) test applied to US real GDP growth, as described in Hansen (1992). The sample starts is 1947:Q2. The specification is $y_t = \mu + \rho_1 y_{t-1} + \rho_2 y_{t-1} + \sigma\epsilon_t$, where y_t is real GDP growth. For each parameter of the specification, the null hypothesis is that the respective parameter is constant.

Table B.1:
TEST RESULTS OF NYBLOM TEST

	L_c	
	AR(1)	AR(2)
μ	0.518*	0.473*
ρ_1	0.367	0.331
ρ_2		0.094
σ^2	0.843***	0.838***
Joint L_c	2.145***	2.294***

Notes: Results are obtained using Nyblom's L test as described in Hansen (1992). *, ** and *** indicate significance at the 10%, 5% and 1% level.

B.2 Bai and Perron Test

Table B.2 reports the result for the Bai and Perron (1998) test applied to US real GDP growth for the sample starting in 1947:Q2. We apply the $SupF_T(k)$ test for the null hypothesis of no break against the alternatives of $k = 1, 2$, or 3 breaks. Secondly, the test $SupF_T(k+1|k)$ tests the null of k breaks against the alternative of $k+1$ breaks. Finally, the $U_d max$ statistic tests the null of absence of break against the alternative of an unknown number of breaks. The null hypothesis of no breaks is rejected against the alternative of one break at the 10% level. The null is not rejected against the alternative of two or three breaks. Furthermore, the null hypothesis of one break against two breaks, or the null of only two against three breaks is not rejected. The final test confirms the conclusion that there is some evidence in favor of at least one break, with the null rejected against an unknown number of breaks at the 10% level. The most likely break is identified to have happened in the second quarter of 2000.

Table B.2:
TEST RESULTS OF BAI-PERRON TEST

	Sample 1947-2015
	$SupF_T(k)$
$k = 1$	8.379*
	[2000:Q2]
$k = 2$	4.194
	[1968:Q2; 2000:Q2]
$k = 3$	4.337
	[1969:Q1; 1982:Q4; 2000:Q2]
	$SupF_T(k\|k-1)$
$k = 2$	1.109
$k = 3$	2.398
$U_d max$	8.379*

Note: Results are obtained using the Bai and Perron (1998) methodology. Dates in square brackets are the most likely break date(s) for each of the specifications. * indicates significance at the 10% level.

C Monte Carlo Evidence

C.1 Setup for Monte Carlo simulations

To assess the performance of our model in the presence of potentially relevant types of misspecification, we carry out a variety of Monte Carlo experiments. In each experiment, we simulate a large number of data sets which are generated from the model under known parameter values, and estimate our model repeatedly over these data sets. This appendix presents the results for two sets of such experiments, which are designed to explore the robustness of crucial assumptions made in the paper.

- In Section C.2 we examine whether the random walk assumption for the time-varying parameters is robust to a different type of structural change. In particular, we verify how the model performs if the underlying long-run growth rate of GDP features one or multiple discrete breaks rather than gradual change. We also estimate our baseline model on data which is generated with a constant instead of a time-varying long-run growth rate of real GDP growth. Furthermore, we repeat this type of experiment for discrete breaks rather than gradual change in the volatilities of both the common factor and the idiosyncratic terms.

- In Section C.3 we explore the robustness of our model to the presence of (unmodeled) change in the long-run growth rate of other series. We entertain the possibility that such unmodeled trends are either independent of the change in the long-run growth real GDP growth or that some series share the trend of GDP. We also verify robustness to both of these types of misspecification simultaneously.

While our simulations feature selected types of misspecification, we aim to ensure a realistic environment for the correctly specified parts of the model. In particular, we set the values of the parameters to their estimated posterior median of the US results. We then take draws for the random disturbances and generate a sample of the vector of 28 observables using equations (1) to (7), and generate 800 periods of data, which corresponds to the monthly sample size in our US application.[1] The four quarterly series are generated by simulating the underlying monthly series and then introducing missing observations by (backwards) applying the polynomial in equation (9). We then estimate the model using the settings described in the paper. The number of simulations (repeatedly drawn samples) per given experiment is set to 100.[2]

[1]While we argue in the paper that the random walk assumption for the estimation of the time-varying parameters is innocuous, it can be problematic to simulate data from parameters that follow random walks. Although we would like the parameters to drift in a non-stationary fashion, i.e. to generate realistic patterns of time-varying volatility, data sets generated from "explosive" processes feature unrealistic properties. To address this issue in the Monte Carlo simulations we discard and re-generate random walks when they drift across a fix threshold. For example, we do not allow the range of (demeaned) time-varying intercept of a given series to exceed the range of its cyclical component.

[2]In certain cases, convergence of the algorithm takes longer in the presence of misspecification, which required us to increase the number of draws of the Gibbs sampler, and thus limited the amount of repetitions that was feasible for a given experiment.

C.2 Results: Sensitivity of random walk specification

The goal of this first set of Monte Carlo experiments is to explore the sensitivity of our modeling choice with respect to the random walk specification of the time-varying parameters. The details about how we justify this modeling assumption can be found in Section 3.1 of the paper. In particular, we aim here to verify whether the model is robust in a context in which there are changes in the long-run growth rate of real GDP growth and in the volatility of business cycles, but these changes occur as discrete breaks rather then as gradual change. Figures C.1 to C.4 present the results of four Monte Carlo experiments.

In the first experiment, the simulated counterpart of real GDP growth features a mean growth rate that is constant but subject to a level shift in the middle of the sample. In Figure C.1, panels (a) and (b), we plot the actual growth rate underlying the data-generating process together with one and two standard deviation percentiles of the 100 simulations of the posterior median, both for the filtered and smoothed estimate. It is reassuring to see that the random walk process "learns" relatively quickly about the underlying change, even in the case of a discrete jump. Panel (c) displays the true, together with the posterior estimate of the common factor for one of the 100 Monte Carlo draws. Panel (d) provides a scatter plot of the true vs. estimated stochastic volatilities. Both pictures show that the models performs well at capturing the simulated objects.

In the second experiment, we repeat the same exercise in the presence of two discrete breaks in the real GDP growth rate. The results are visible in Figure C.2, which tells a very similar story to the first experiment. We omit panels for factor and the stochastic volatility estimates, as they are very similar to the first experiment.

In the third experiment, we verify the consequences of estimating our model in an in environment in which the parameters which we specify as time-varying are in fact constant in the data-generating process. The results, displayed in Figure C.3, confirm that the random walk assumption appears to be entirely innocuous in this setting. Both the long-run GDP growth rate (smoothed and filtered), as well as the volatility of the factor are estimated to be constant, with relatively high precision. In addition, similar to the first experiment, the estimate of the common factor is very precise.

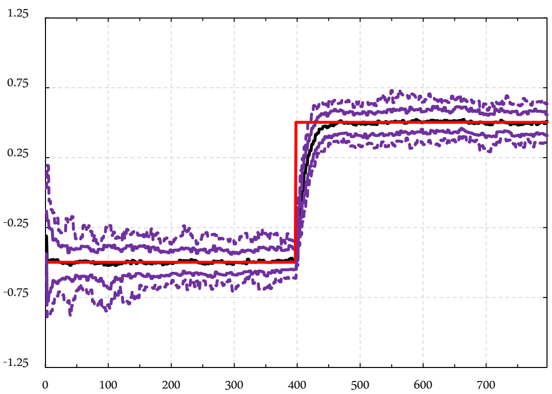
Finally, in the fourth experiment, we again keep the long-run growth rate of real GDP constant but this time introduce a discrete shift in the volatilities of both the common factor and the idiosyncratic terms of all series in the middle of generated data sample. Reassuringly, the shift in the volatilities is well captures in the estimation and does not spill over to the estimate of the long-run growth rate of real GDP.

In conclusion from these experiments, the random walk assumption appears to be flexible enough to accommodate structural change that occurs in discrete steps rather than gradually. This underpins our conclusions about the apparent gradual changes in the long-run growth of the US economy described in the paper.

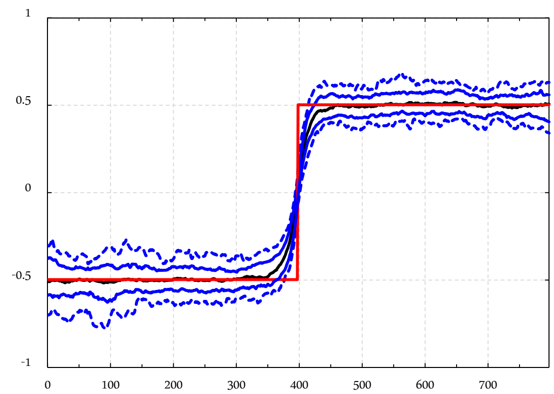
Figure C.1: Simulation Results I

Data-generating process (DGP) with one discrete break in long-run real GDP growth

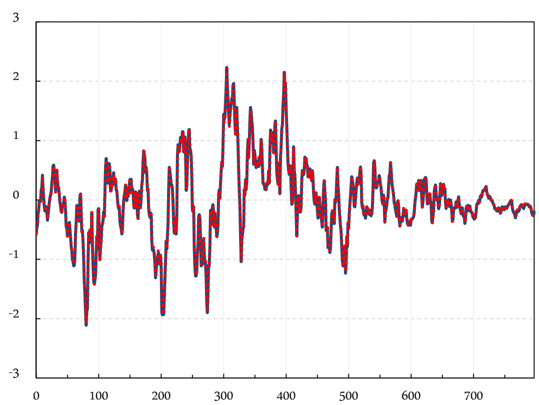
(a) True vs. Estimated Trend (Filtered)



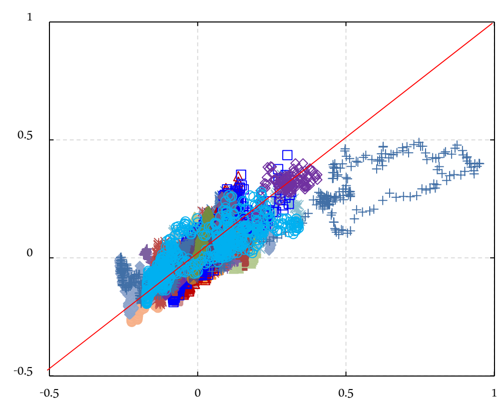
(b) True vs. Estimated Trend (Smoothed)



(c) True vs. Estimated Factor



(d) True vs. Estimated Volatilities of $\mathbf{u}_t$

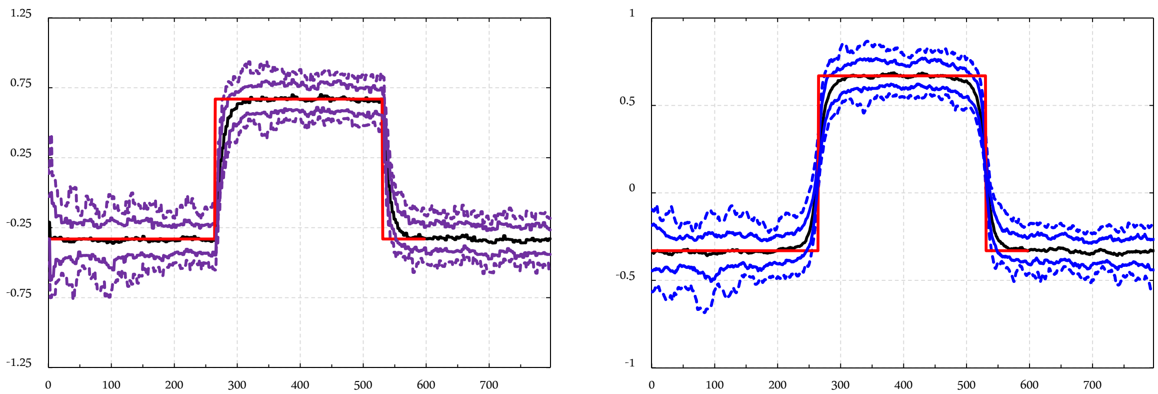


Note: The DGP features a discrete break in the trend of GDP growth occurring in the middle of the sample, as well as stochastic volatility. The sample size is $n = 28$ and $T = 800$, which mimics our US data set. The estimation procedure is the fully specified model as defined by equations (1)-(7) in the text. We carry out 100 simulations drawing from the DGP. Panel (a) presents the long-run growth component as estimated by the Kalman filter, plotted against the actual long-run growth rate generated from the DGP. The corresponding figure for the smoothed estimate is given in panel (b). In both panels, the median (black) as well the 68th (solid) and 90th (dashed) percentile of the 100 simulated outcomes are shown in blue/purple. Panel (c) displays the factor generated by the the DGP (red) and its smoothed estimate (blue) for one draw. Panel (d) provides evidence on the accuracy of the estimation of the SV of the idiosyncratic terms, by plotting the volatilities from the DGP against the estimates for the 24 monthly indicators. Both are normalized by subtracting the average volatility.

Figure C.2: Simulation Results II

Data-generating process (DGP) with two discrete breaks in in long-run real GDP growth

(a) True vs. Estimated Trend (Filtered) (b) True vs. Estimated Trend (Smoothed)

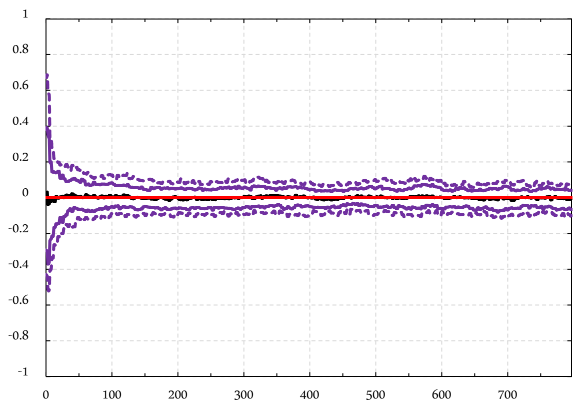


Note: The simulation setup is equivalent to the one in Figure C.1 but features *two* discrete breaks in the trend at 1/3 and 2/3 of the sample. Again, we show the filtered as well as the smoothed trend median estimates and the corresponding 68th and 90th percentiles of the 100 simulated estimates of these objects. Panels (c) and (d) are omitted as they are very similar to Figure C.1.

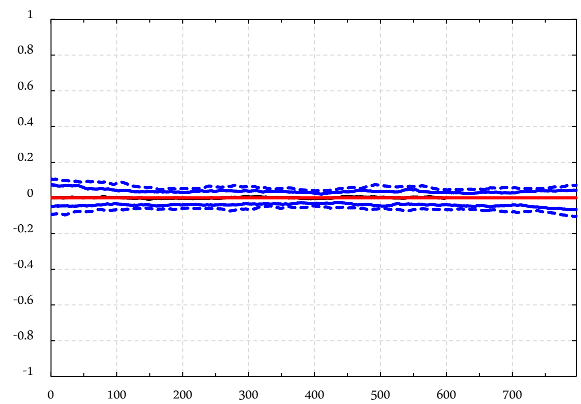
Figure C.3: Simulation Results III

Data-generating process (DGP) without in changes long-run real GDP growth and without SV

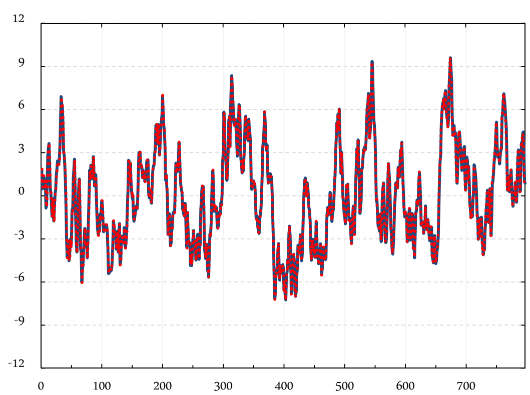
(a) True vs. Estimated Trend (Filtered)



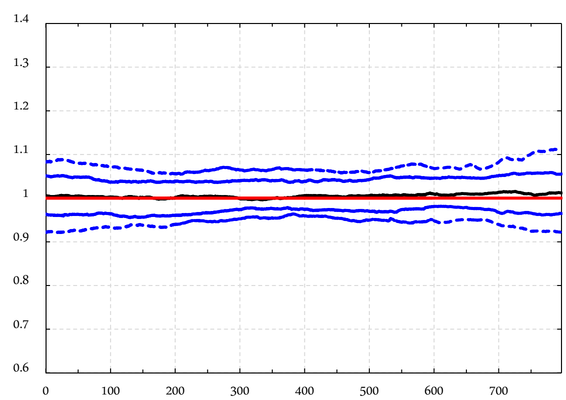
(b) True vs. Estimated Trend (Smoothed)



(c) True vs. Estimated Factor



(d) True vs. Estimated Volatility of Factor

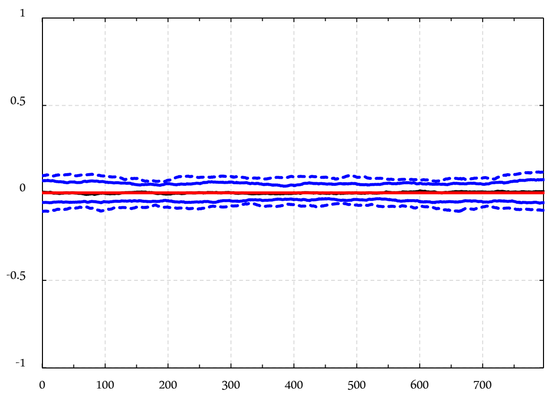


Note: The DGP is the baseline model without trend in GDP growth and without stochastic volatility. The estimation procedure is the fully specified model as explained in the description of Figure C.1. Again, we plot the filtered and smoothed median estimates of the long-run growth rate with 68th and 90th percentiles of the 100 simulated estimates in panels (a) and (b). Panel (c) presents a comparison of the estimated factor and its DGP counterpart for one Monte Carlo draw. Panel (d) in similar to (b), but for the volatility of the common factor.

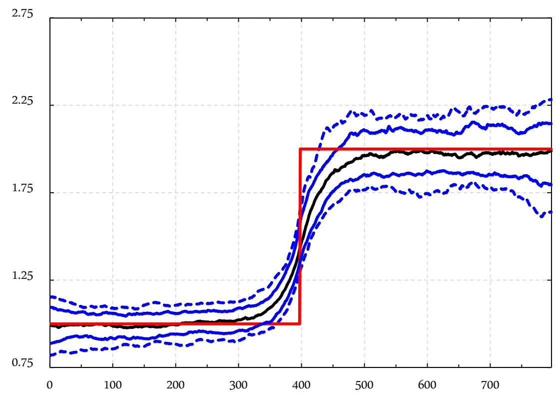
Figure C.4: Simulation Results IV

Data-generating process (DGP) with discrete break in volatility

(a) True vs. Estimated Trend (Smoothed)



(b) True vs. Estimated Volatility of Factor



Note: The DGP does not feature any changes in the trend of GDP growth, but one discrete break in the volatility of the common factor. As in Figures C.1-C.3, the estimation procedure is based on the fully specified mode. Panel (a) displays the smoothed posterior median estimate of the trend component of GDP growth, with 68th and 90th percentiles of the 100 simulations shown as solid and dashed blue lines, respectively. Panel (b) displays the posterior median estimate of the volatility of the common factor (black), with the corresponding percentiles.

C.3 Results: Sensitivity to confounding time-variation

Our model can flexibly accommodate time-varying intercepts in all or a subset of the series contained in our data panel. Given our interest in tracking real GDP growth, we restrict our baseline model to feature a trend in GDP only (shared by consumption) and argue that such unmodeled time-variation is picked up by the idiosyncratic components, which we allow to be persistent. Details about this discussion are contained in Section 3.2 of the paper. The goal of this second set of Monte Carlo experiments is to verify how robust our model is in a setting where time-varying intercepts are indeed present in the data-generating process but not modeled explicitly in the estimation. Figures C.5 to C.7 present the results of three Monte Carlo experiments in which such "confounding trends" are added when generating the data.[3]

In the first experiment, the misspecification arises from the fact that our model explicitly specifies a time-varying mean in the GDP equation only, while the data is generated such that the first 18 series of the panel all feature independent non-stationary means.[4] Figure C.5 presents the estimation results in this setup. Panel (a) shows the percentiles of the deviations of the estimated from the actual real GDP growth rates over the 100 simulations (repeated draws from the DGP). The percentiles are centered relatively tightly around zero, meaning that the trend estimates with 68 and 90% smallest deviations are relatively similar to the original trend process. To illustrate this further, panels (b), (c) and (d) display more detailed results for one of the 100 Monte Carlo simulations, labeled "Median Simulation". This is selected by ordering the outcomes of all repeated samples by the distance of squared deviation of the estimated from the simulated GDP trend and then selecting the median. This essentially means that 49% of the simulations had larger, and 50% smaller deviations than the simulation displayed. The panels plot actual against estimated (black/red) long-run real GDP growth rate, common factor and factor volatility, respectively. In the case of the long-run growth rate the posterior credible intervals are added in blue. These results reveal that in a typical (median) outcome for this type of specification, the model performs well at capturing these objects. Most importantly, the "true" long-run growth rate is contained within the posterior bands throughout the entire estimation sample.

In the second experiment, the data-generating process features a single time-varying mean which is present in the first 6 series, whereas we still only specify it in the first series for the estimation.[5] The results for this experiment are shown in Figure C.6. The panels here are similar to Figure C.5. While the deviations in panel (a) are slightly larger than for the previous figure, indicating that common unmodeled trends are somewhat more challenging to pick up than independent ones, the overall message remains the qualitatively similar. In particular, the results for the "Median Experiment", displayed in panels (b) to (d), are reassuring in that the estimate tracks their data counterpart closely.

[3] For simplicity we assume that the estimated model in this section is the one with a trend in GDP only, i.e. $\mathbf{B} = 1$.

[4] Formally, in the DGP $dim(\mathbf{a}_t) = 18$ and $\mathbf{B} = \mathbf{I}_{18}$, while the model for estimation is specified by $\mathbf{a}_t = g_t$ and $\mathbf{B} = 1$. We assume that the remaining 10 of the 28 series are stationary, which mimics the presence of the surveys in our data set.

[5] In our notation this means that in the DGP we have $\mathbf{a}_t = g_t$ and $\mathbf{B} = \mathbf{1}_{6\times 1}$, while the model for estimation is specified by $\mathbf{a}_t = g_t$ and $\mathbf{B} = 1$. We choose 6 series so that both quarterly and monthly variables are affected by the misspecification.

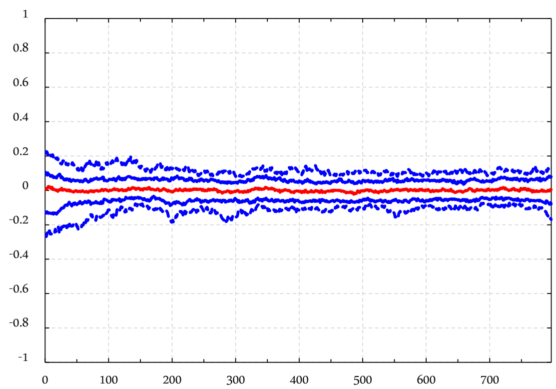
The third experiment introduces both types of misspecification simultaneously, i.e. independent time-varying means in series 1-18 and an additional shared time-varying component in series 1-6. The results are presented in Figure C.7. The take-aways are similar to the previous figures, even in the presence of this heavy type of misspecification.

Overall, these simulation experiments confirm our intuition that the estimate of the time-varying mean of interest is not affected by low frequency movements present in other series that are not explicitly modeled. Despite the extremely unfavorable assumption of a large amount of additional time-variation, the long-run growth rate of real GDP is tracked very well in all settings considered.

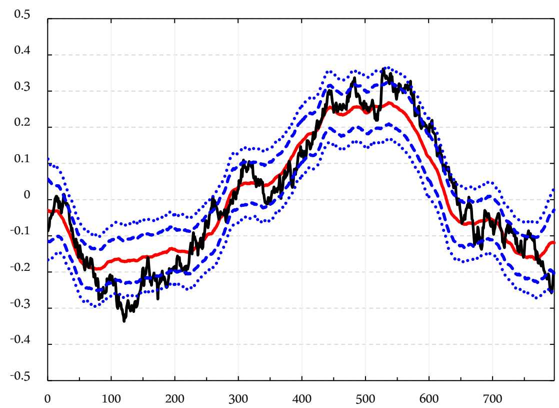
Figure C.5: Simulation Results V

Data-generating process (DGP) with independent unmodeled trends in other series

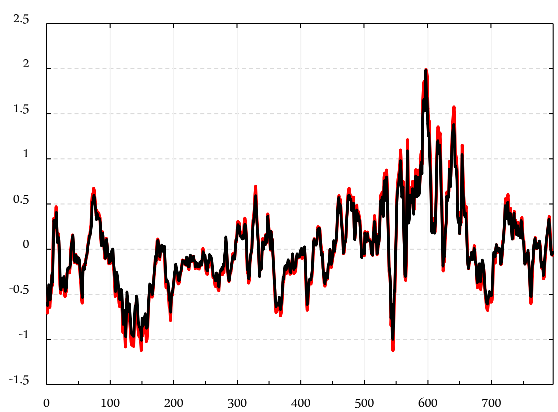
(a) True vs. Est. Trend - Deviation Percentiles



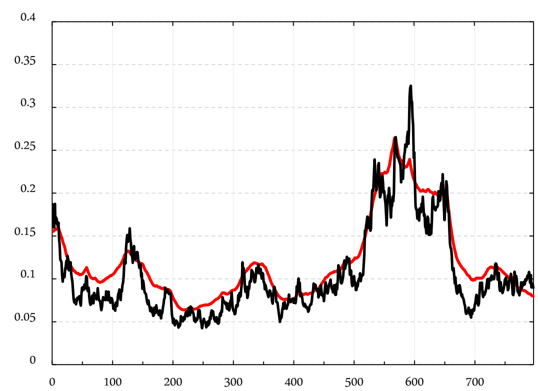
(b) True vs. Est. Trend (Median Simulation)



(c) True vs. Est. Factor (Median Simulation)



(d) True vs. Est. Vol (Median Simulation)

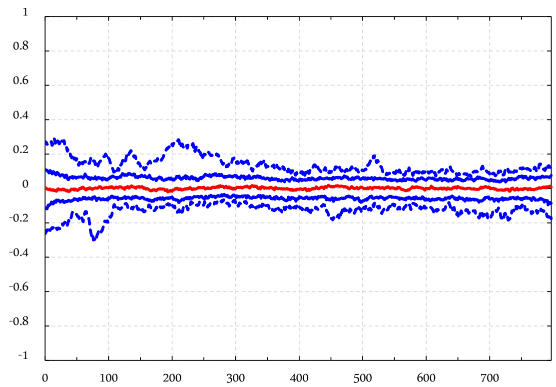


Note: The DGP features independent time-varying means in series 1-18. The sample size is $n = 28$ and $T = 800$, which mimics our US data set. The estimation procedure is the fully specified model as defined by equations (1)-(7) in the text, with a time-varying mean only specified for the real GDP growth equation. We carry out a Monte Carlo simulation with 100 samples repeatedly drawn from the DGP. Panel (a) presents the median (red), as well as the 68 and 90% bands (blue) of the deviation of the estimated long-run growth rate from its actual data counterpart over 100 simulated outcomes. Panel (b) shows the true (black) together with the posterior median estimate (red) of the long-run growth rate of real GDP. The 68% (solid blue) and 90% (dashed blue) posterior credible intervals are also plotted. Panels (c) and (d) plot the median estimate (red) against true (black) common factor and its stochastic volatility.

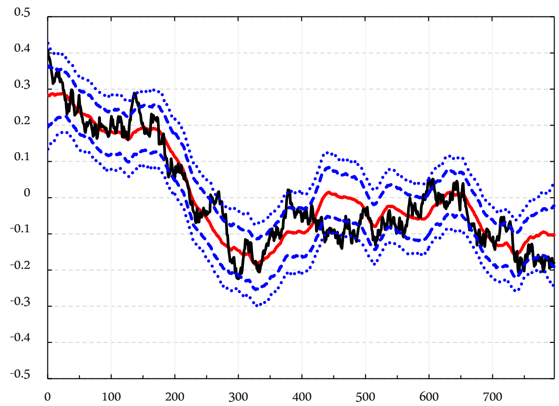
Figure C.6: Simulation Results VI

Data-generating process (DGP) with shared unmodeled trends in other series

(a) True vs. Est. Trend - Deviation Percentiles



(b) True vs. Est. Trend (Median Simulation)



(c) True vs. Est. Factor (Median Simulation)

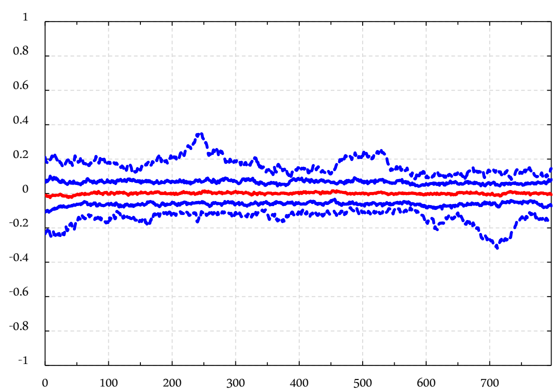
(d) True vs. Est. Vol (Median Simulation)

Note: The DGP features a common time-varying in series 1-6, while the estimation specifies this stochastic trend only in the equation for real GDP growth. The rest of the setup of the simulations, as well as the structure of the panels are similar to Figure C.5.

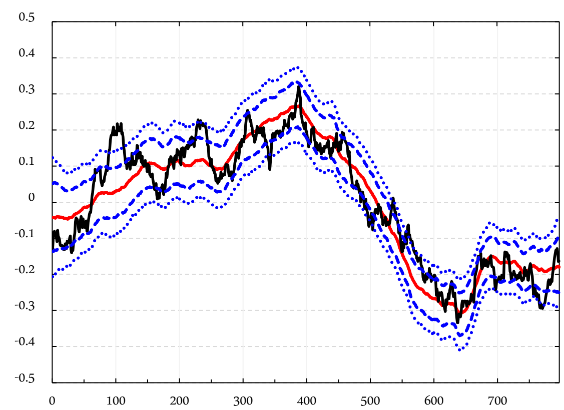
Figure C.7: Simulation Results VII

Data-generating process (DGP) with both independent and shared unmodeled trends in other series

(a) True vs. Est. Trend - Deviation Percentiles



(b) True vs. Est. Trend (Median Simulation)



(c) True vs. Est. Factor (Median Simulation)

(d) True vs. Est. Vol (Median Simulation)

The DGP features both independent time-varying components in series 1-18 as well as a common time-varying in series 1-6, while the estimation specifies a stochastic trend only in the equation for real GDP growth. The rest of the setup of the simulations, as well as the structure of the panels are similar to Figure C.5

D Details on Estimation Procedure

D.1 Construction of the State Space System

For expositional clarity, we focus on the baseline case with $\mathbf{B} = 1$ and $\mathbf{a}_t = a_t$ here, so that $m = r = 1$. Recall that in our main specification we choose the order of the polynomials in equations (3) and (4) to be $p = 2$ and $q = 2$, respectively. Let the $n \times 1$ vector $\tilde{\mathbf{y}}_t$, which contains n_q de-meaned quarterly and n_m de-meaned monthly variables (i.e. $n = n_q + n_m$), be defined as

$$\tilde{\mathbf{y}}_t = \begin{bmatrix} y^q_{1,t} \\ \vdots \\ y^q_{n_q,t} \\ y^m_{1,t} - \rho^m_{1,1} y^m_{1,t-1} - \rho^m_{1,2} y^m_{1,t-2} \\ \vdots \\ y^m_{n_m,t} - \rho^m_{n_m,1} y^m_{n_m,t-1} - \rho^m_{n_m,2} y^m_{n_m,t-2} \end{bmatrix},$$

so that the system is written out in terms of the *quasi-differences* of the monthly indicators. Given this re-defined vector of observables, we cast our model into the following state space form:

$$\begin{aligned} \tilde{\mathbf{y}}_\mathbf{t} &= \mathbf{H}\mathbf{X}_t + \tilde{\boldsymbol{\eta}}_\mathbf{t}, && \tilde{\boldsymbol{\eta}}_\mathbf{t} \sim N(0, \tilde{\mathbf{R}}_\mathbf{t}) \\ \mathbf{X}_t &= \mathbf{F}\mathbf{X}_{t-1} + \mathbf{e}_t, && \mathbf{e}_t \sim N(0, \mathbf{Q}_t) \end{aligned}$$

where the state vector is defined as $\mathbf{X}'_t = [a_t, \dots, a_{t-4}, f_t, \dots, f_{t-4}, \mathbf{u}^{q\prime}_t, \dots, \mathbf{u}^{q}_{t-4}{}']$. Setting $\lambda_1 = 1$ for identification, the matrices of parameters $\mathbf{H}$ and $\mathbf{F}$, are then constructed as shown below:

$$\mathbf{H} = \left[\begin{array}{c|c|c} \mathbf{H}_a & \begin{matrix} \mathbf{H}_{\lambda_q} \\ \mathbf{H}_{\lambda_m} \end{matrix} & \mathbf{H}_u \end{array} \right],$$

where the respective blocks of $\mathbf{H}$ are defined as

$$\mathbf{H}_a = \left[\begin{array}{ccccc} \frac{1}{3} & \frac{2}{3} & 1 & \frac{2}{3} & \frac{1}{3} \\ \hline & & \mathbf{0}_{(n-1)\times 5} & & \end{array} \right], \qquad \mathbf{H}_{\lambda_q} = \begin{bmatrix} 1 & \lambda_2 & \dots & \lambda_{n_q} \end{bmatrix}' \times \begin{bmatrix} 1/3 & 2/3 & 1 & 2/3 & 1/3 \end{bmatrix},$$

$$\mathbf{H}_{\lambda_m} = \left[\begin{array}{c|c} \lambda_{n_q+1} - \lambda_{n_q+1}\rho^m_{1,1} - \lambda_{n_q+1}\rho^m_{1,2} & \mathbf{0}_{1\times 4} \\ \vdots & \vdots \\ \lambda_n - \lambda_n \rho^m_{n_m,1} - \lambda_n \rho^m_{n_m,2} & \mathbf{0}_{1\times 4} \end{array} \right]$$

$$\mathbf{H}_u = \begin{bmatrix} \bar{\mathbf{H}}_u \\ \mathbf{0}_{n_m \times 5} \end{bmatrix}, \qquad \bar{\mathbf{H}}_u = \mathbf{1}_{n_q \times 1} \times \begin{bmatrix} 1/3 & 2/3 & 1 & 2/3 & 1/3 \end{bmatrix},$$

and

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_1 & \mathbf{0} & & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2 & & & \\ \vdots & & \mathbf{F}_{2+1} & & \vdots \\ \vdots & & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \dots & \mathbf{0} & \mathbf{F}_{2+n_q} \end{bmatrix},$$

where the respective blocks of $\mathbf{F}$ are defined as

$$\mathbf{F}_1 = \left[\begin{array}{cc} 1 & \mathbf{0}_{1\times 4} \\ \hline \mathbf{I}_4 & \mathbf{0}_{4\times 1} \end{array}\right] \qquad \mathbf{F}_2 = \left[\begin{array}{ccc} \phi_1 & \phi_2 & \mathbf{0}_{1\times 3} \\ \hline \multicolumn{2}{c}{\mathbf{I}_4} & \mathbf{0}_{4\times 1} \end{array}\right] \qquad \mathbf{F}_{2+j} = \left[\begin{array}{ccc} \rho^q_{j,1} & \rho^q_{j,2} & \mathbf{0}_{1\times 3} \\ \hline \multicolumn{2}{c}{\mathbf{I}_4} & \mathbf{0}_{4\times 1} \end{array}\right]$$

for $j = 1, ..., n_q$.

The error terms are denoted as

$$\begin{aligned} \tilde{\boldsymbol{\eta}}_t &= [\mathbf{0}_{1\times n_q}, \tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}^{\boldsymbol{m}\prime}]' \\ \mathbf{e}_t &= \begin{bmatrix} v_{a_t} & \mathbf{0}_{4\times 1} & \epsilon_t & \mathbf{0}_{4\times 1} & \eta_{1,t} & \mathbf{0}_{4\times 1} & \dots & \eta_{n_q,t} & \mathbf{0}_{4\times 1} \end{bmatrix}', \end{aligned}$$

with covariance matrices

$$\tilde{\mathbf{R}}_{\mathbf{t}} = \begin{bmatrix} \mathbf{0}_{n_q\times n_q} & \mathbf{0}_{n_q\times n_m} \\ \mathbf{0}_{n_m\times n_q} & \mathbf{R}_t \end{bmatrix},$$

where $\mathbf{R}_t = diag(\sigma^2_{\eta^m_{1,t}}, \dots, \sigma^2_{\eta^m_{n_m,t}})$ and

$$\mathbf{Q}_t = diag(\omega^2_a, \mathbf{0}_{1\times 4}, \sigma^2_{\epsilon,t}, \mathbf{0}_{1\times 4}, \sigma^2_{\eta^q_{1,t}}, \mathbf{0}_{1\times 4}, \dots, \sigma^2_{\eta^q_{n_q,t}}, \mathbf{0}_{1\times 4}).$$

D.2 Details of the Gibbs Sampler

For ease of notation, we again restrict this description to the case of one time-varying mean specified as $m = r = 1$, $\mathbf{B} = 1$ and $\mathbf{a}_t = a_t$. Let $\boldsymbol{\theta} \equiv \{\boldsymbol{\lambda}, \boldsymbol{\Phi}, \boldsymbol{\rho}, \omega_a, \omega_\varepsilon, \omega_{\eta_1}, \ldots, \omega_{\eta_n}\}$ be a vector that collects the underlying parameters, where $\boldsymbol{\Phi}$ and $\boldsymbol{\rho}$ contain the parameters for factor and idiosyncratic components respectively. The model is estimated using a Markov Chain Monte Carlo (MCMC) Gibbs sampling algorithm in which conditional draws of the latent variables, $\{a_t, f_t\}_{t=1}^T$, the parameters, $\boldsymbol{\theta}$, and the stochastic volatilities, $\{\sigma_{\varepsilon,t}, \sigma_{\eta_{i,t}}\}_{t=1}^T$ are obtained sequentially. The algorithm has a block structure composed of the following steps.

0. Initialization

The model parameters are initialized at arbitrary starting values $\boldsymbol{\theta}^0$, and so are the sequences for the stochastic volatilities, $\{\sigma^0_{\varepsilon,t}, \sigma^0_{\eta_{i,t}}\}_{t=1}^T$. Set $j = 1$.

1. Draw latent variables conditional on model parameters and SVs

Obtain a draw $\{a_t^j, f_t^j, \mathbf{u}_t^q\}_{t=1}^T$ from $p(\{a_t, f_t\}_{t=1}^T | \boldsymbol{\theta}^{j-1}, \{\sigma^{j-1}_{\varepsilon,t}, \sigma^{j-1}_{\eta_{i,t}}\}_{t=1}^T, \mathbf{y})$.

This step of the algorithm uses the state space representation described above (Appendix D.1), and produces a draw from the entire state vector $\mathbf{X}_t$ (which includes the long-run growth components, a_t, the common factor, f_t, and the idiosyncratic components of the quarterly variables, $\mathbf{u}_t^q$) by means of a forward-filtering backward-smoothing algorithm, see Carter and Kohn (1994) or Kim and Nelson (1999b). In particular, we adapt the algorithm proposed by Bai and Wang (2015), which is robust to numerical inaccuracies, and extend it to the case with mixed frequencies and missing data following Mariano and Murasawa (2003), as explained in section 3.3. Like Bai and Wang (2015), we initialise the Kalman Filter step from a normal distribution whose moments are independent of the model parameters, in particular $\mathbf{X}_0 \sim N(0, 10^4\mathbf{I})$.

2. Draw the variance of the time-varying GDP growth component

Obtain a draw $\omega_a^{2,j}$ from $p(\omega_a^2 | \{a_t^j\}_{t=1}^T)$.

Taking the sample $\{a_t^j\}_{t=1}^T$ drawn in the previous step as given, and posing an inverse-gamma prior $p(\omega_a^2) \sim IG(S_a, v_a)$ the conditional posterior of ω_a^2 is also drawn inverse-gamma distribution. As discussed in Section 4.2, we choose the scale $S_a = 10^{-3}$ and degrees of freedom $v_a = 1$ for our baseline specification.

3. Draw the autoregressive parameters of the factor VAR

Obtain a draw $\boldsymbol{\Phi}^j$ from $p(\boldsymbol{\Phi} | \{f_t^{j-1}, \sigma^{j-1}_{\varepsilon,t}\}_{t=1}^T)$.

Taking the sequences of the common factor $\{f_t^{j-1}\}_{t=1}^T$ and its stochastic volatility $\{\sigma^{j-1}_{\varepsilon,t}\}_{t=1}^T$ from previous steps as given, and posing a non-informative prior, the corresponding conditional posterior is drawn from the Normal distribution, see, e.g. Kim and Nelson

(1999b). In the more general case of more than one factor, this step would be equivalent to drawing from the coefficients of a Bayesian VAR. Like Kim and Nelson (1999b), or Cogley and Sargent (2005), we reject draws which imply autoregressive coefficients in the explosive region.

4. Draw the factor loadings

Obtain a draw of $\boldsymbol{\lambda}^j$ from $p(\boldsymbol{\lambda}|\rho^{j-1}, \{f_t^{j-1}, \sigma_{\eta_{i,t}}^{j-1}\}_{t=1}^T, \mathbf{y})$.

Conditional on the draw of the common factor $\{f_t^{j-1}\}_{t=1}^T$, the measurement equations reduce to n independent linear regressions with heteroskedastic and serially correlated residuals. By conditioning on $\boldsymbol{\rho}^{j-1}$ and $\sigma_{\eta_{i,t}}^{j-1}$, the loadings can be estimated using GLS and non-informative priors. When necessary, we apply restrictions on the loadings using the formulas provided by de Wind and Gambetti (2014), see Appendix D.3 for further information.

5. Draw the serial correlation coefficients of the idiosyncratic components

Obtain a draw of $\boldsymbol{\rho}^j$ from $p(\boldsymbol{\rho}|\lambda^{j-1}, \{f_t^{j-1}, \sigma_{\eta_{i,t}}^{j-1}\}_{t=1}^T, \mathbf{y})$.

Taking the sequence of the common factor $\{f_t^{j-1}\}_{t=1}^T$ and the loadings drawn in previous steps as given, the idiosyncratic components for the monthly variables can be obtained as $u_{i,t} = y_{i,t} - \lambda^{j-1} f_t^{j-1}$. For the quarterly variables, a draw of the idiosyncratic components has been obtained directly from Step 1. Given a sequence for the stochastic volatility of the i^{th} component, $\{\sigma_{\eta_{i,t}}^{j-1}\}_{t=1}^T$, the residual is standardized to obtain an autoregression with homoskedastic residuals whose conditional posterior can be drawn from the Normal distribution as in step 2.3.

6. Draw the stochastic volatilities

Obtain a draw of $\{\sigma_{\varepsilon,t}^j\}_{t=1}^T$ and $\{\sigma_{\eta_{i,t}}^j\}_{t=1}^T$ from $p(\{\sigma_{\varepsilon,t}\}_{t=1}^T|\boldsymbol{\Phi}^{j-1}, \{f_t^{j-1}\}_{t=1}^T)$, and from $p(\{\sigma_{\eta_{i,t}}\}_{t=1}^T|\boldsymbol{\lambda}^{j-1}, \boldsymbol{\rho}^{j-1}, \{f_t^{j-1}\}_{t=1}^T, \mathbf{y})$ respectively.

Finally, we draw the stochastic volatilities of the innovations to the factor and the idiosyncratic components independently, using the algorithm proposed by Kim et al. (1998), which uses a mixture of normal random variables to approximate the elements of the log-variance. This is a more efficient alternative to the exact Metropolis-Hastings algorithm previously proposed by Jacquier et al. (2002). For the general case in which there is more than one factor, the volatilities of the factor VAR can be drawn jointly, see Primiceri (2005).

Increase j by 1, go to Step 2.1 and iterate until convergence is achieved.

D.3 Implementing linear restrictions on the factor loadings

To impose linear restrictions on the factor loadings $\boldsymbol{\lambda}$ in equation (1) of the paper, we follow de Wind and Gambetti (2014). For linear restrictions of the form

$$\mathbf{R}\boldsymbol{\lambda} = \mathbf{r} \tag{11}$$

these authors consider the special case with $\mathbf{r} = \mathbf{0}$ in equation (54) in the appendix to their paper. For $\mathbf{r} \neq \mathbf{0}$, this equation is amended as shown here. Let $\boldsymbol{\lambda}^u$ and $\boldsymbol{\lambda}^r$ denote the unrestricted and restricted loading matrix, respectively. $\boldsymbol{\lambda}^r$ is then drawn from a posterior distribution defined by (12) to (14):

$$\boldsymbol{\lambda}^r \sim N\left(\overline{\boldsymbol{\lambda}}^r, \mathbf{P}_\lambda^r\right), \tag{12}$$

where

$$\overline{\boldsymbol{\lambda}}^r = \boldsymbol{\lambda}^u - \mathbf{P}_\lambda^u \mathbf{R}' \left(\mathbf{R}\mathbf{P}_\lambda^u \mathbf{R}'\right)^{-1} \left(\mathbf{R}\boldsymbol{\lambda}^u - \mathbf{r}\right) \tag{13}$$

$$\mathbf{P}_\lambda^r = \mathbf{P}_\lambda^u - \mathbf{P}_\lambda^u \mathbf{R}' \left(\mathbf{R}\mathbf{P}_\lambda^u \mathbf{R}'\right)^{-1} \mathbf{R}\mathbf{P}_\lambda^u. \tag{14}$$

E Details on the Construction of the Data Base

E.1 US (Vintage) Data Base

For our US real-time forecasting evaluation, we consider data vintages since 11 January 2000 capturing the real activity variables listed in the text. For each vintage, the start of the sample is set to January 1960, appending missing observations to any series which starts after that date. All times series are obtained from one of these sources: (1) Archival Federal Reserve Economic Data (ALFRED), (2) Bloomberg, (3) Haver Analytics. Table E.1 provides details on each series, including the variable code corresponding to the different sources.

For several series, in particular Retail Sales, New Orders, Imports and Exports, only vintages in nominal terms are available, but series for appropriate deflators are available from Haver, and these are not subject to revisions. We therefore deflate them using, respectively, CPI, PPI for Capital Equipment, and Imports and Exports price indices. Additionally, in several occasions the series for New Orders, Personal Consumption, Vehicle Sales and Retail Sales are subject to methodological changes and part of their history gets discontinued. In this case, given our interest in using long samples for all series, we use older vintages to splice the growth rates back to the earliest possible date.

For *soft* variables real-time data is not as readily available. The literature on real-time forecasting has generally assumed that these series are unrevised, and therefore used the latest available vintage. However while the underlying survey responses are indeed not revised, the seasonal adjustment procedures applied to them do lead to important differences between the series as was available at the time and the latest vintage. For this reason we use seasonally un-adjusted data and re-apply the Census-X12 procedure in real time to obtain a real-time seasonally adjusted version of the surveys. We follow the same procedure for the initial unemployment claims series. We then use Bloomberg to obtain the exact date in which each monthly data point was first published.

Table E.1:
DETAILED DESCRIPTION OF DATA SERIES

	Frequ.	Start Date	Vintage Start	Trans-formation	Publ. Lag	Data Code
Real Gross Domestic Product	Q	Q2:1947	Dec 91	%QoQ Ann	26	GDPC1(F)
Real Consumption (ex. durables)	Q	Q2:1947	Dec 91	%QoQ Ann	26	
Hours worked	Q	Q2:1948	Dec 91	%QoQ Ann	28	
Real Investment (incl. durable cons.)	Q	Q2:1947	Dec 91	%QoQ Ann	26	
Real Industrial Production	M	Jan 47	Jan 97	% MoM	15	INDPRO(F)
Real Manufacturers' New Orders Nondefense Capital Goods Excluding Aircraft	M	Mar 68	Mar 97	% MoM	25	NEWORDER(F)[1] PPICPE(F)
Real Light Weight Vehicle Sales	M	Feb 67	Mar 97	% MoM	1	ALTSALES(F)[2] TLVAR(H)
Real Personal Income less Transfer Payments	M	Feb 59	Dec 97	% MoM	27	DSPIC96(F)
Real Retail Sales Food Services	M	Feb 47	Jun 01	% MoM	15	RETAIL(F) CPIAUCSL(F) RRSFS(F)[3]
Real Exports of Goods	M	Feb 68	Jan 97	% MoM	35	BOPGEXP(F)[4] C111CPX(H) TMXA(H)
Real Imports of Goods	M	Feb 69	Jan 97	% MoM	35	BOPGIMP(F)[4] C111CP(H) TMMCA(H)
Building Permits	M	Feb 60	Aug 99	% MoM	19	PERMIT(F)
Housing Starts	M	Feb 59	Jul 70	% MoM	26	HOUST(F)
New Home Sales	M	Feb 63	Jul 99	% MoM	26	HSN1F(F)
Total Nonfarm Payroll Employment (Establishment Survey)	M	Jan 47	May 55	% MoM	5	PAYEMS(F)
Civilian Employment (Household Survey)	M	Feb 48	Feb 61	% MoM	5	CE16OV(F)
Unemployed	M	Feb 48	Feb 61	% MoM	5	UNEMPLOY(F)
Initial Claims for UE	M	Feb 48	Jan 00*	% MoM	4	LICM(H)

(Continues on next page)

Detailed Description of Data Series (Continued)

Markit Manufacturing PMI	M	May 07	Jan 00*	-	-7	S111VPMM(H)[5] H111VPMM(H)
ISM Manufacturing PMI	M	Jan 48	Jan 00*	-	1	NMFBAI(H) NMFNI(H) NMFEI(H) NMFVDI(H)[6]
ISM Non-manufacturing PMI	M	Jul 97	Jan 00*	-	3	NAPMCN(H)
Conference Board: Consumer Confidence	M	Feb 68	Jan 00*	Diff 12 M.	-5	CCIN(H)
University of Michigan: Consumer Sentiment	M	May 60	Jan 00*	Diff 12 M.	-15	CSENT(H)[5] CONSSENT(F) Index(B)
Richmond Fed Manufacturing Survey	M	Nov 93	Jan 00*	-	-5	RIMSXN(H) RIMNXN(H) RIMLXN(H)[6]
Philadelphia Fed Business Outlook	M	May 68	Jan 00*	-	0	BOCNOIN(H) BOCNONN(H) BOCSHNN(H) BOCDTIN(H) BOCNENN(H)[6]
Chicago PMI	M	Feb 67	Jan 00*	-	0	PMCXPD(H) PMCXNO(H) PMCXI(H) PMCXVD(H)[6]
NFIB: Small Business Optimism Index	M	Oct 75	Jan 00*	Diff 12 M.	15	NFIBBN (H)
Empire State Manufacturing Survey	M	Jul 01	Jan 00*	-	-15	EMNHN(H) EMSHN(H) EMDHN(H) EMDSN(H) EMESN(H)[6]

Notes: The second column refers to the sampling frequency of the data, which can be quarterly (Q) or monthly (M). % QoQ Ann. refers to the quarter on quarter annualized growth rate, % MoM refers to $(y_t - y_{t-1})/y_{t-1}$ while Diff 12 M. refers to $y_t - y_{t-12}$. In the last column, (B) = Bloomberg; (F) = FRED; (H) = Haver;
1) deflated using PPI for capital equipment; 2) for historical data not available in ALFRED we used data coming from HAVER; 3) using deflated nominal series up to May 2001 and real series afterwards; 4) nominal series from ALFRED and price indices from HAVER. For historical data not available in ALFRED we used data coming from HAVER; 5) preliminary series considered; 6) NSA subcomponents needed to compute the SA headline index. * Denotes seasonally un-adjusted series which have been seasonally adjusted in real time.

E.2 Data Base for Other G7 Economies

Table E.2:
Canada

	Freq.	Start Date	Transformation
Real Gross Domestic Product	Q	Jun-1960	% QoQ Ann.
Industrial Production: Manuf., Mining, Util.	M	Jan-1960	% MoM
Manufacturing New Orders	M	Feb-1960	% MoM
Manufacturing Turnover	M	Feb-1960	% MoM
New Passenger Car Sales	M	Jan-1960	% MoM
Real Retail Sales	M	Feb-1970	% MoM
Construction: Dwellings Started	M	Feb-1960	% MoM
Residential Building Permits Auth.	M	Jan-1960	% MoM
Real Exports	M	Jan-1960	% MoM
Real Imports	M	Jan-1960	% MoM
Unemployment Ins.: Initial and Renewal Claims	M	Jan-1960	% MoM
Employment: Industrial Aggr. excl. Unclassified	M	Feb-1991	% MoM
Employment: Both Sexes, 15 Years and Over	M	Feb-1960	% MoM
Unemployment: Both Sexes, 15 Years and Over	M	Feb-1960	% MoM
Consumer Confidence Indicator	M	Jan-1981	Diff 12 M.
Ivey Purchasing Managers Index	M	Jan-2001	Level
ISM Manufacturing PMI	M	Jan-1960	Level
University of Michigan: Consumer Sentiment	M	May-1960	Diff 12 M.

Notes: The second column refers to the sampling frequency of the data, which can be quarterly (Q) or monthly (M). % QoQ Ann. refers to the quarter on quarter annualized growth rate, % MoM refers to $(y_t - y_{t-1})/y_{t-1}$ while Diff 12 M. refers to $y_t - y_{t-12}$. All series were obtained from the Haver Analytics database.

Table E.3:
GERMANY

	Freq.	Start Date	Transformation
Real Gross Domestic Product	Q	Jun-1960	% QoQ Ann.
Mfg Survey: Production: Future Tendency	M	Jan-1960	Level
Ifo Demand vs. Prev. Month: Manufact.	M	Jan-1961	Level
Ifo Business Expectations: All Sectors	M	Jan-1991	Level
Markit Manufacturing PMI	M	Apr-1996	Level
Markit Services PMI	M	Jun-1997	Level
Industrial Production	M	Jan-1960	% MoM
Manufacturing Turnover	M	Feb-1960	% MoM
Manufacturing Orders	M	Jan-1960	% MoM
New Truck Registrations	M	Feb-1963	% MoM
Total Unemployed	M	Feb-1962	% MoM
Total Domestic Employment	M	Feb-1981	% MoM
Job Vacancies	M	Feb-1960	% MoM
Retail Sales Volume excluding Motor Vehicles	M	Jan-1960	% MoM
Wholesale Vol. excl. Motor Veh. and Motorcycles	M	Feb-1994	% MoM
Real Exports of Goods	M	Feb-1970	% MoM
Real Imports of Goods	M	Feb-1970	% MoM

Notes: The second column refers to the sampling frequency of the data, which can be quarterly (Q) or monthly (M). % QoQ Ann. refers to the quarter on quarter annualized growth rate, % MoM refers to $(y_t - y_{t-1})/y_{t-1}$ while Diff 12 M. refers to $y_t - y_{t-12}$. All series were obtained from the Haver Analytics database.

Table E.4:
JAPAN

	Freq.	Start Date	Transformation
Real Gross Domestic Product	Q	Jun-1960	% QoQ Ann.
TANKAN All Industries: Actual Business Cond.	Q	Sep-1974	Diff 1 M.
Markit Manufacturing PMI	M	Oct-2001	Level
Small Business Sales Forecast	M	Dec-1974	Level
Small/Medium Business Survey	M	Apr-1976	Level
Consumer Confidence Index	M	Mar-1973	Level
Inventory to Sales Ratio	M	Jan-1978	Level
Industrial Production: Mining and Manufact.	M	Jan-1960	% MoM
Electric Power Consumed by Large Users	M	Feb-1960	% MoM
New Motor Vehicle Registration: Trucks, Total	M	Feb-1965	Diff 1 M.
New Motor Vehicle Reg: Passenger Cars	M	May-1968	% MoM
Real Retail Sales	M	Feb-1960	% MoM
Real Department Store Sales	M	Feb-1970	% MoM
Real Wholesale Sales: Total	M	Aug-1978	% MoM
Tertiary Industry Activity Index	M	Feb-1988	% MoM
Labor Force Survey: Total Unemployed	M	Jan-1960	% MoM
Overtime Hours / Total Hours (manufact.)	M	Feb-1990	% MoM
New Job Offers excl. New Graduates	M	Feb-1963	% MoM
Ratio of New Job Openings to Applications	M	Feb-1963	% MoM
Ratio of Active Job Openings and Active Job Appl.	M	Feb-1963	% MoM
Building Starts, Floor Area: Total	M	Feb-1965	% MoM
Housing Starts: New Construction	M	Feb-1960	% MoM
Real Exports	M	Feb-1960	% MoM
Real Imports	M	Feb-1960	% MoM

Notes: The second column refers to the sampling frequency of the data, which can be quarterly (Q) or monthly (M). % QoQ Ann. refers to the quarter on quarter annualized growth rate, % MoM refers to $(y_t - y_{t-1})/y_{t-1}$ while Diff 12 M. refers to $y_t - y_{t-12}$. All series were obtained from the Haver Analytics database.

Table E.5:
United Kingdom

	Freq.	Start Date	Transformation
Real Gross Domestic Product	Q	Mar-1960	% QoQ Ann.
Dist. Trades: Total Vol. of Sales	M	Jul-1983	Level
Dist. Trades: Retail Vol. of Sales	M	Jul-1983	Leve
CBI Industrial Trends: Vol. of Output Next 3 M.	M	Feb-1975	Level
BoE Agents' Survey: Cons. Services Turnover	M	Jul-1997	Level
Markit Manufacturing PMI	M	Jan-1992	Level
Markit Services PMI	M	Jul-1996	Level
Markit Construction PMI	M	Apr-1997	Level
GfK Consumer Confidence Barometer	M	Jan-1975	Diff 12 M.
Industrial Production: Manufacturing	M	Jan-1960	% MoM
Passenger Car Registrations	M	Jan-1960	% MoM
Retail Sales Volume: All Retail incl. Autom. Fuel	M	Jan-1960	% MoM
Index of Services: Total Service Industries	M	Feb-1997	% MoM
Registered Unemployment	M	Feb-1960	% MoM
Job Vacancies	M	Feb-1960	% MoM
LFS: Unemployed: Aged 16 and Over	M	Mar-1971	% MoM
LFS: Employment: Aged 16 and Over	M	Mar-1971	% MoM
Mortgage Loans Approved: All Lenders	M	May-1993	% MoM
Real Exports	M	Feb-1961	% MoM
Real Imports	M	Feb-1961	% MoM

Notes: The second column refers to the sampling frequency of the data, which can be quarterly (Q) or monthly (M). % QoQ Ann. refers to the quarter on quarter annualized growth rate, % MoM refers to $(y_t - y_{t-1})/y_{t-1}$ while Diff 12 M. refers to $y_t - y_{t-12}$. All series were obtained from the Haver Analytics database.

Table E.6:
FRANCE

	Freq.	Start Date	Transformation
Real Gross Domestic Product	Q	Jun-1960	% QoQ Ann.
Industrial Production	M	Feb-1960	% MoM
Total Commercial Vehicle Registrations	M	Feb-1975	% MoM
Household Consumption Exp.: Durable Goods	M	Feb-1980	% MoM
Real Retail Sales	M	Feb-1975	% MoM
Passenger Cars	M	Feb-1960	% MoM
Job Vacancies	M	Feb-1989	% MoM
Registered Unemployment	M	Feb-1960	% MoM
Housing Permits	M	Feb-1960	% MoM
Housing Starts	M	Feb-1974	% MoM
Volume of Imports	M	Jan-1960	% MoM
Volume of Exports	M	Jan-1960	% MoM
Business Survey: Personal Prod. Expect.	M	Jun-1962	Level
Business Survey: Recent Output Changes	M	Jan-1966	Level
Household Survey: Household Conf. Indicator	M	Oct-1973	Diff 12 M.
BdF Bus. Survey: Production vs. Last M., Ind.	M	Jan-1976	Level
BdF Bus. Survey: Production Forecast, Ind.	M	Jan-1976	Level
BdF Bus. Survey: Total Orders vs. Last M., Ind.	M	Jan-1981	Level
BdF Bus. Survey: Activity vs. Last M., Services	M	Oct-2002	Level
BdF Bus. Survey: Activity Forecast, Services	M	Oct-2002	Level
Markit Manufacturing PMI	M	Apr-1998	Level
Markit Services PMI	M	May-1998	Level

Notes: The second column refers to the sampling frequency of the data, which can be quarterly (Q) or monthly (M). % QoQ Ann. refers to the quarter on quarter annualized growth rate, % MoM refers to $(y_t - y_{t-1})/y_{t-1}$ while Diff 12 M. refers to $y_t - y_{t-12}$. All series were obtained from the Haver Analytics database.

Table E.7:
ITALY

	Freq.	Start Date	Transformation
Real Gross Domestic Product	Q	Jun-1960	% QoQ Ann.
Markit Manufacturing PMI	M	Jun-1997	Level
Markit Services PMI: Business Activity	M	Jan-1998	Level
Production Future Tendency	M	Jan-1962	Level
ISTAT Services Survey: Orders, Next 3 M-	M	Jan-2003	Level
ISTAT Retail Trade Confidence Indicator	M	Jan-1990	Level
Industrial Production	M	Jan-1960	% MoM
Real Exports	M	Jan-1960	% MoM
Real Imports	M	Jan-1960	% MoM
Real Retail Sales	M	Feb-1990	% MoM
Passenger Car Registrations	M	Jan-1960	% MoM
Employed	M	Feb-2004	% MoM
Unemployed	M	Feb-1983	% MoM

Notes: The second column refers to the sampling frequency of the data, which can be quarterly (Q) or monthly (M). % QoQ Ann. refers to the quarter on quarter annualized growth rate, % MoM refers to $(y_t - y_{t-1})/y_{t-1}$ while Diff 12 M. refers to $y_t - y_{t-12}$. All series were obtained from the Haver Analytics database.

F Choice of Priors

As explained in the paper, we use non-informative priors for loadings and serial correlation coefficients of factor and idiosyncratic components in order to aide comparability with the previous literature, which has generally used classical estimation methods. With respect to the choice of priors on the new parameters of our specification, namely ω_a^2, ω_ε^2 and $\omega_{\eta,i}^2$ in equations (5)-(7), we closely follow the related literature, in particular Cogley and Sargent (2005) and Primiceri (2005), by setting relatively conservative priors, which shrink the model towards the benchmark with no time-variation, but are still loose enough for the data to be able to speak. In particular, in all the inverse-gamma (IG) distributions we set the number of degrees of freedom to 1, the minimum required to make the prior distributions proper while keeping the weight of the prior low. As to the choice of the scale parameter of the IG distributions, it is worth pointing out that this does not parametrize time variation itself, but rather incorporates a prior belief about the amount of time variation. To gain an intuition about the prior on ω_a^2, in Section 4.2 we note that the chosen value of 0.001 implies that over a period of ten years the random walk process of the long-run growth rate is expected to vary with a standard deviation of around 0.4 percentage points in annualized terms, which we believe is a fairly conservative prior in terms of economic magnitudes. The choice of 10^{-4} for the prior on ω_ε^2 and $\omega_{\eta,i}^2$ is similar to the approach of Primiceri (2005).

To shed some light on the robustness of our results to the choice of priors, in what follows we explore the sensitivity of our main results to varying the tightness of the respective priors. To summarize the most notable finding, we find that the data strongly drives the result of time variation both in the long-run growth rate and the volatilies: a dogmatically large amount of shrinkage is needed in order to make either of them disappear.

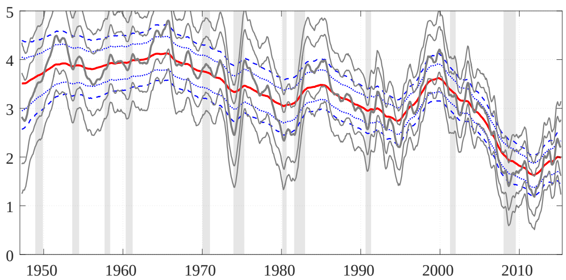
F.1 Robustness checks on prior choice

Prior on innovation variance to the time-varying long-run growth rate

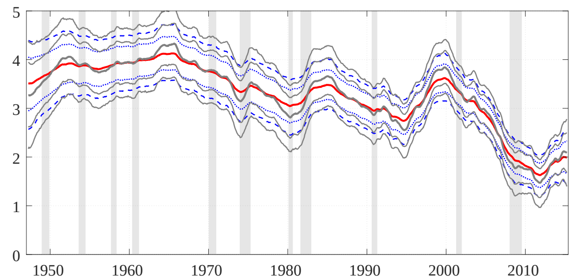
In Figure F.1 we explore the sensitivity of our key results to the choice of the scale parameter of the prior on the innovation variance to the time-varying long-run growth rate of real GDP, ω_a^2. Each panel plots our baseline estimate of g_t, which has been obtained with a prior scale of 10^{-3} (red/blue). We then successively compare this baseline estimate with alternative estimates obtained when imposing both looser and tighter prior scales, respectively (gray). Panel (a) of the figure reveals that with a prior implying a very large variance the estimated trend is pinned down with relatively more uncertainty and evolves in bumpy fashion, yet the qualitative pattern around the evolution of long-run growth, in particular the recent slowdown, remains clearly visible. Panels (b) and (c) show that using a ten times looser prior (0.01) and a hundred times tighter prior (10^{-5}) than the one in our baseline setting gives very similar results to ours. In the later case, the estimate is almost identical. Finally, a dogmatically tight prior (10^{-9}) is required to make variation in the long-run growth rate disappear entirely, which is visible in Panel (d).

Figure F.1: Comparison Across Different Prior Scales of ω_α^2

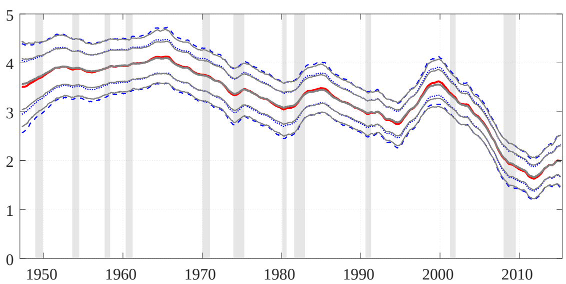
(a) $\omega_a^2 = 0.1$ (loose prior)



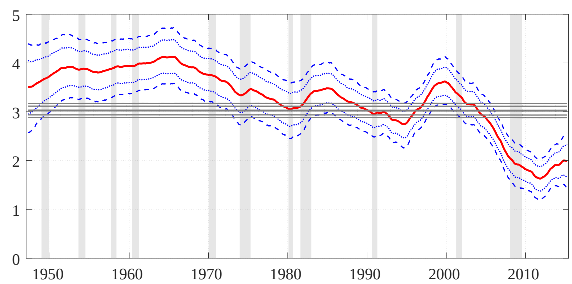
(b) $\omega_a^2 = 0.01$ (looser prior)



(c) $\omega_a^2 = 10^{-5}$ (very tight prior)



(d) $\omega_a^2 = 10^{-9}$ (extremely tight prior)



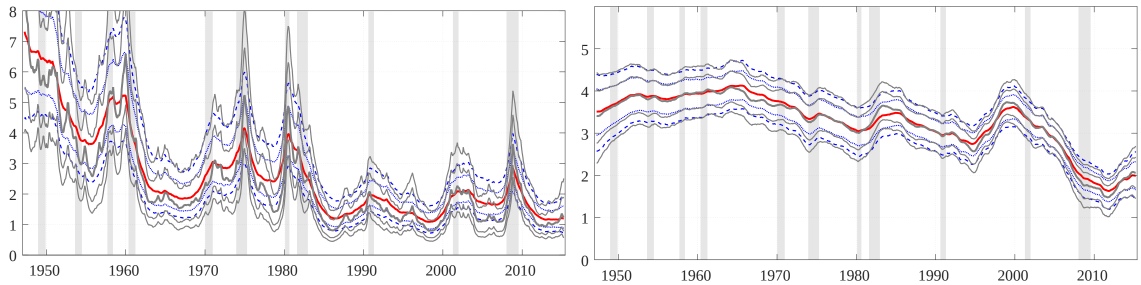
Note: In each panel our baseline the median estimate of real GDP growth based on a scale of 10^{-3} is presented (red), with corresponding 68% (solid blue) and the 90% (dashed blue) posterior credible intervals. The corresponding estimates based an different prior scales are superimposed in gray in each panel.

Prior on innovation variance to the SV

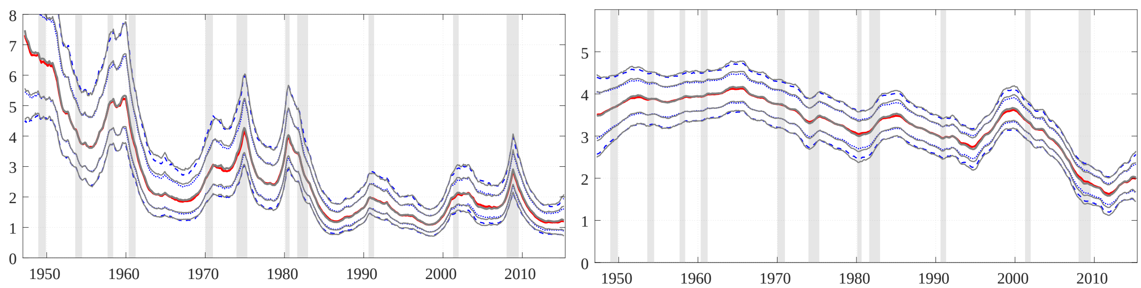
Figure F.2 presents the sensitivity of the results to the choice of the scale parameter of the prior on the innovation variance to the SV in both the common factor and the idiosyncratic components. Similar to Figure F.1 we compare our baseline estimates (red/blue), where we set $\omega_\varepsilon^2 = \omega_{\eta,i}^2 = 10^{-4}$, with estimates obtained under a range of varied prior scales (gray). In each case, the figure shows both the estimated SV of the factor as well as the estimate of the long-run growth rate of real GDP growth. Panel (a) displays the results for a very loose prior (1), while Panel (b) for a prior which is ten times looser than the baseline (10^{-3}). Finally, the estimates shown in Panel (c) are obtained under a tighter prior (10^{-5}). Again, the results reported in the paper do not seem to be affected. Both the estimates of the SV and the long-run growth rate of real GDP are almost identical to our main results.

Figure F.2: Comparison Across Different Prior Scales of ω_{ε}^2 and $\omega_{\eta,i}^2$

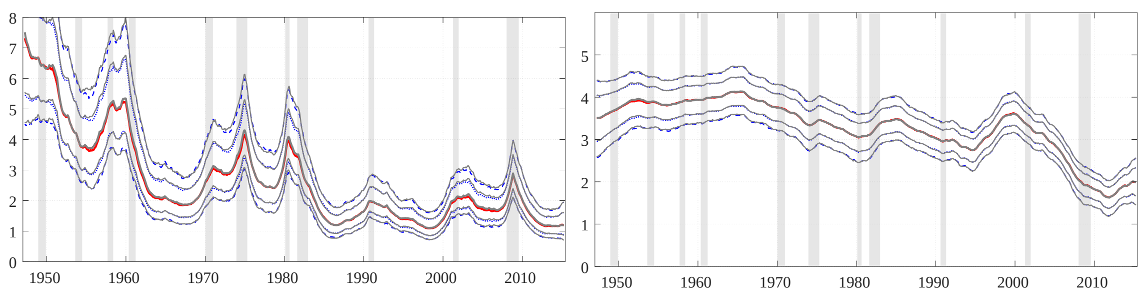
(a) $\omega_{\varepsilon}^2 = \omega_{\eta,i}^2 = 1$ (very loose prior)



(b) $\omega_{\varepsilon}^2 = \omega_{\eta,i}^2 = 10^{-3}$ (looser prior)



(c) $\omega_{\varepsilon}^2 = \omega_{\eta,i}^2 = 10^{-5}$ (tighter prior)



Note: In each panel our baseline estimate of the SV of the common factor based on a scale of 10^{-4} is presented (red) in the left chart. The right chart plots the estimate of the long-run growth rate of real GDP based on the same scale. Corresponding 68% (solid blue) and the 90% (dashed blue) posterior credible intervals are also plotted. The analogue estimates based on the alternative prior scales are superimposed in gray in each panel.

Prior on serial correlation in factor and idiosyncratic components

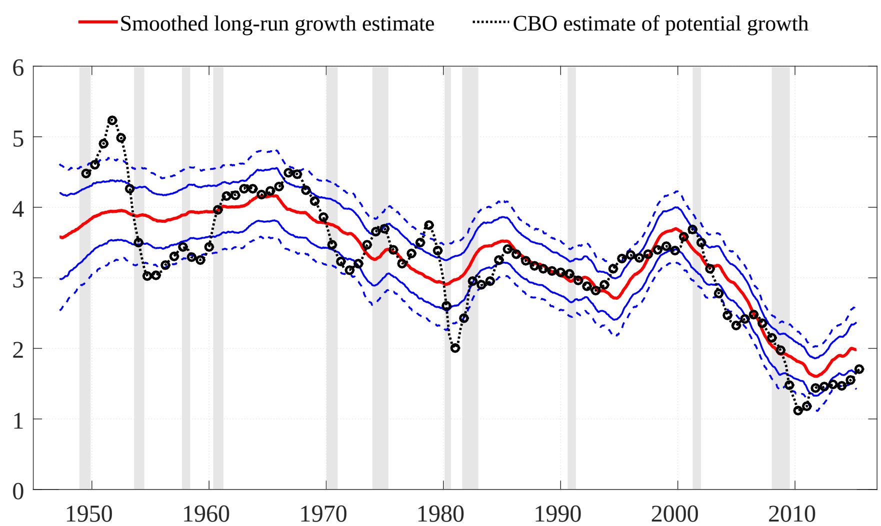
As a final robustness check, we consider "Minnesota"-style priors on the autoregressive coefficients of the factor as well as shrinking the coefficients of the serial correlation towards zero. To be precise, we center the prior on the first lag of the factor around 0.9 and all other lags at zero. The motivation for these priors is to express a preference for a more parsimonious model where the factors capture the bulk of the persistence of the series and the idiosyncratic components are close to iid, that is closer to true measurement error. These alternative priors do not meaningfully affect the posterior estimates of our main objects of interest, so we omit additional figures. Note that we have found some evidence that the use of such priors might at times improve the convergence of the algorithm. Specifically, when we apply the model to the other G7 economies (see Section 5), we find that for some countries where few monthly indicators are available, shrinking the serial correlations of the idiosyncratic components towards zero helps obtaining a common factor that is persistent.

G Comparison with CBO Measure of Potential Output

Our estimate of long-run growth and the CBO's potential growth estimate capture related but not identical concepts. The CBO measures the growth rate of potential output, i.e. the level of output that could be obtained if all resources were used fully, whereas our estimate, similar to Beveridge and Nelson (1981), measures the component of the growth rate that is expected to be permanent. Moreover, the CBO estimate is constructed using the so-called "production function approach", which is radically different from the DFM methodology.[6]

As a simple sanity check, it is interesting to see that despite employing different statistical methods they produce qualitatively similar results, visible in Figure G.1, with the CBO estimate displaying a more marked cyclical pattern but remaining for most of the sample within the 90% credible posterior interval of our estimate. As in our estimate, most of the slowdown occurred prior to the Great Recession. The CBO's estimate was below ours immediately after the recession, reaching an unprecedented low level of about 1.25% in 2010, and remains in the lower bound of our posterior estimate since then. Section 4.6 expands on the reason for this divergence and argues that this is likely to stem from the larger amount of information incorporated in the DFM. In fact, the CBO estimate of potential growth is noticeably more cyclical.

Figure G.1: Long-run GDP Growth Estimate in Comparison to CBO



Note: The figure displays the posterior median estimate of long-run GDP growth with the corresponding credible intervals, as displayed in Figure 2 Panel (a) in the main body of the paper, in comparison with the CBO's measure of potential output growth, which is shown in black circles.

[6]Essentially, the production function approach calculates the trend components of the supply inputs to a neoclassical production function (the capital stock, total factor productivity, and the total amount of hours) using statistical filters and then aggregates them to obtain an estimate of the trend level of output. See CBO (2001).

H Details About the Forecast Evaluation

H.1 Setup

Using our real-time database of US vintages, we re-estimate the following three models each day in which new data is released: a benchmark with constant long-run GDP growth and constant volatility (Model 0, similar to Banbura and Modugno (2014)), a version with constant long-run growth but with stochastic volatility (Model 1, similar to Marcellino et al. (2014)), and the baseline model put forward in the paper with both time-variation in the long-run growth of real GDP and SV (Model 2). Allowing for an intermediate benchmark with only SV allows us to evaluate how much of the improvement in the model can be attributed to the addition of the long-run variation in GDP as opposed to the SV. We evaluate the point and density forecast accuracy relative to the initial ("Advance") release of GDP, which is released between 25 and 30 days after the end of the reference quarter.[7]

When comparing the three different models, we test the significance of any improvement of Models 1 and 2 relative to Model 0. This raises some important econometric complications given that (i) the three models are nested, (ii) the forecasts are produced using an expanding window, and (iii) the data used is subject to revision. These three issues imply that commonly used test statistics for forecasting accuracy, such as the one proposed by Diebold and Mariano (1995) and Giacomini and White (2006) will have a non-standard limiting distribution. However, rather than not reporting any test, we follow the "pragmatic approach" of Faust and Wright (2013) and Groen et al. (2013), who build on Monte Carlo results in Clark and McCracken (2012). Their results indicate that the Harvey et al. (1997) small sample correction of the Diebold and Mariano (1995) statistic results in a good sized test of the null hypothesis of equal finite sample forecast precision for both nested and non-nested models, including cases with expanded window-based model updating. Overall, the results of the tests should be interpreted more as a rough gauge of the significance of the improvement than a definitive answer to the question. We compute various point and density forecast accuracy measures at different moments in the release calendar, to assess how the arrival of information improves the performance of the model. In particular, the computations are carried out starting 180 days before the end of the reference quarter, and every subsequent day up to 25 days after its end, when the GDP figure for the quarter is usually released. This means that we will evaluate the forecasts of the next quarter, current quarter (nowcast), and the previous quarter (backcast). We consider two different samples for the evaluation: the full sample (2000:Q1-2015:Q1) and the sample covering the recovery since the Great Recession (2009:Q2-2015:Q1).

[7]We have explored the alternative of evaluating the forecasts against subsequent releases, or the latest available vintages. The relative performance of the three models is broadly unchanged, but all models do better at forecasting the initial release. If the objective is to improve the performance of the model relative to the first official release, then ideally an explicit model of the revision process would be desirable. The results are available upon request.

H.2 Point Forecast Evaluation

Figure H.1 shows the results of evaluating the posterior mean as point forecast. We use two criteria, the root mean squared error (RMSE) and the mean absolute error (MAE). As expected, both of these decline as the quarters advance and more information on monthly indicators becomes available, see e.g. Banbura et al. (2012). Both the RMSE and the MAE of Model 2 are lower than that of Model 0, particularly so from the start of the nowcasting period, while Model 1 is somewhat worse overall. Our gauge of significance indicates that these differences in nowcasting performance are significant at the 10% level for the overall sample in the case of the MAE, but not the RMSE. The improvement in performance is much clearer in the recovery sample. In fact, the inclusion of the time varying long run component of GDP helps anchor GDP predictions at a level consistent with the weak recovery experienced in the past few years and produces nowcasts that are 'significantly' superior to those of the reference model from around 30 days before the end of the reference quarter. In essence, ignoring the variation in long-run GDP growth would have resulted in being on average around 1 percentage point too optimistic from 2009 to 2015.

H.3 Density Forecast Evaluation

Density forecasts can be used to assess the ability of a model to predict unusual developments, such as the likelihood of a recession or a strong recovery given current information. The adoption of a Bayesian framework allows us to produce density forecasts from the DFM that consistently incorporate both filtering and estimation uncertainty. Figure H.2 reports the probability integral transform (PITs) and the associated autocorrelation functions (ACFs) for the 3 models calculated with the nowcast of the last day of the quarter. Diebold et al. (1998) highlight that well calibrated densities are associated with uniformly distributed and independent PITs. Figure H.2 suggests that the inclusion of SV is paramount to get well calibrated densities, whereas the inclusion of the long-run growth component helps to get a more appropriate representation of the right side of the distribution, as well as making sure that the first order autocorrelation is not statistically significant.

There are several measures available for density forecast evaluation. The (average) log score, i.e. the logarithm of the predictive density evaluated at the realization, is one of the most popular, rewarding the model that assigns the highest probability to the realized events. Gneiting and Raftery (2007), however, caution against using the log score, emphasizing that it does not appropriately reward values from the predictive density that are close but not equal to the realization, and that it is very sensitive to outliers. They therefore propose the use of the (average) continuous rank probability score (CRPS) in order to address these drawbacks of the log-score. Figure H.3 shows that by both measures our model outperforms its counterparts. Interestingly, the comparison of Model 1 and Model 2 suggests that failing to properly account for the long-run growth component might give a misrepresentation of the GDP densities, resulting in poorer density forecasts.

In addition to the above results, we also assess how the three models fare when different areas of their predictive densities are emphasized in the forecast evaluation. To do that we follow Groen et al. (2013) and compute weighted averages of Gneiting and Raftery (2007) quantile scores (QS) that are based on quantile forecasts that correspond to the predictive densities from the different models (Figure H.4).[8] Our results indicate that while there is an improvement in density nowcasting for the entire distribution, the largest improvement comes from the right tail. For the full sample, Model 1 is very close to Model 0, suggesting that being able to identify the location of the distribution is key to the improvement in performance. In order to appreciate the importance of the improvement in the density forecasts, and in particular in the right side of the distribution, we calculated a recursive estimate of the likelihood of a 'strong recovery', where this is defined as the probability of an average growth rate of GDP (over the present and next three quarters) above the historical average. Model 0 and Model 2 produce very similar probabilities up until 2011 when, thanks to the downward revision of long-run GDP growth, Model 2 starts to deliver lower probability estimates consistent with the observed weak recovery. The Brier score for Model 2 is 0.186 whereas the score for Model 0 is 0.2236 with the difference significantly different at 1% (Model 1 is essentially identical to Model 0).[9]

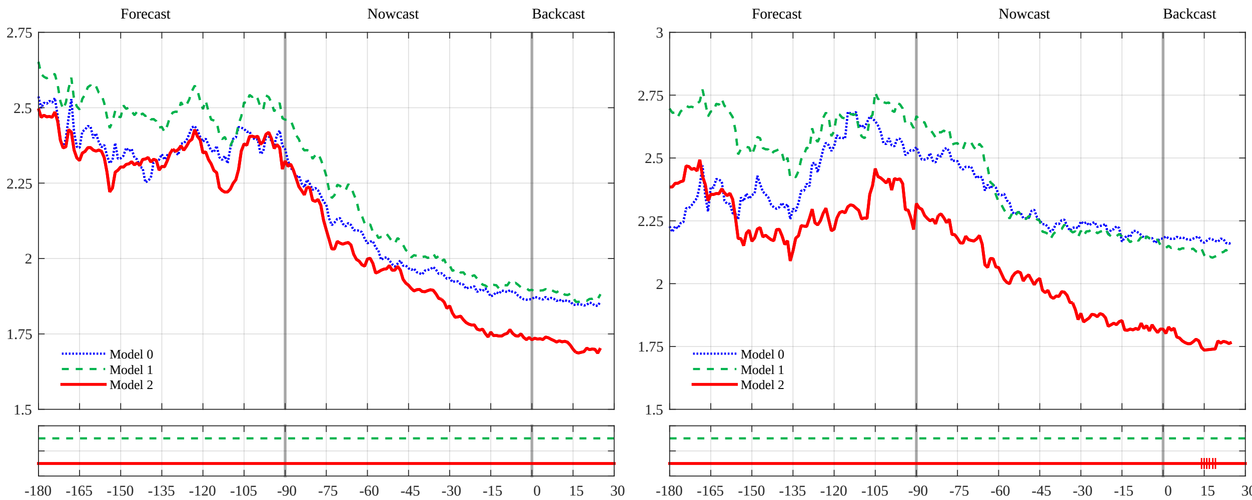
In sum, the results of the out-of-sample forecasting evaluation indicate that a model that allows for time-varying long-run GDP growth and SV produces short-run forecasts that are on average (over the full evaluation sample) either similar to or improve upon the benchmark model. The performance tends to improve substantially in the sub-sample including the recovery from the Great Recession, coinciding with the significant downward revision of the model's assessment of long-run growth. Furthermore, the results indicate that while there is an improvement in density nowcasting for the entire distribution, the largest improvement comes from the right tail.

[8]As Gneiting and Ranjan (2011) show, integrating QS over the quantile spectrum gives the CRPS.
[9]The results are available upon request.

Figure H.1: Point Forecast Accuracy Evaluation

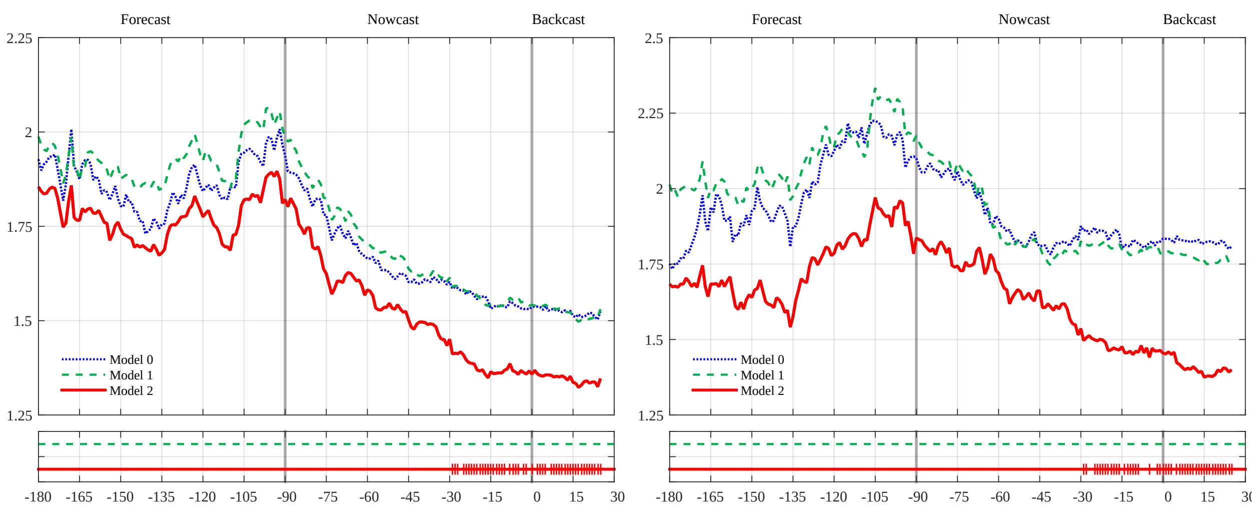
(a) Root Mean Squared Error

Full Sample: 2000:Q1-2015:Q1 Recovery Sample: 2009:Q2-2015:Q1



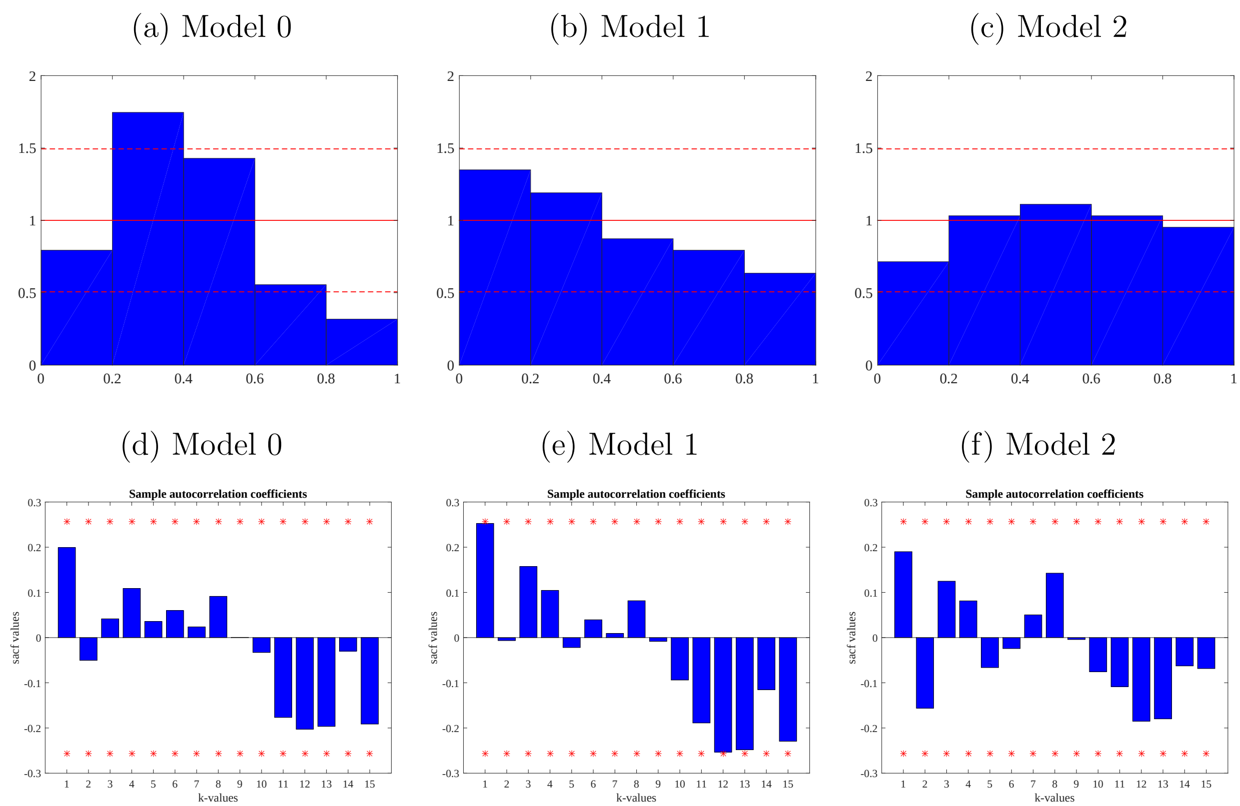
(b) Mean Absolute Error

Full Sample: 2000:Q1-2015:Q1 Recovery Sample: 2009:Q2-2015:Q1



Note: The horizontal axis indicates the forecast horizon, expressed as the number of days to the end of the reference quarter. Thus, from the point of view of the forecaster, forecasts produced 180 to 90 days before the end of a given quarter are a forecast of next quarter; forecasts 90-0 days are nowcasts of current quarter, and the forecasts produced 0-25 days after the end of the quarter are backcasts of last quarter. The boxes below each panel display, with a vertical tick mark, a gauge of statistical significance at the 10% level of any difference with Model 0, for each forecast horizon, as explained in the main text.

Figure H.2: Probability Integral Transform (PITs)

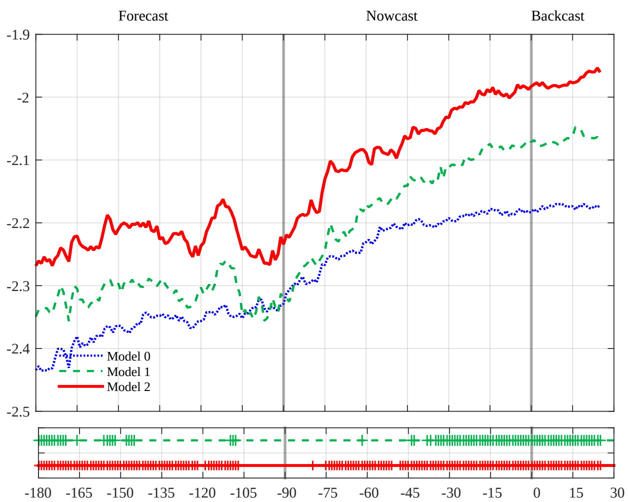


Note: The upper three panels display the cdf of the Probability Integral Transforms (PITs) evaluated on the last day of the reference quarter, while the lower three display the associated autocorrelation functions.

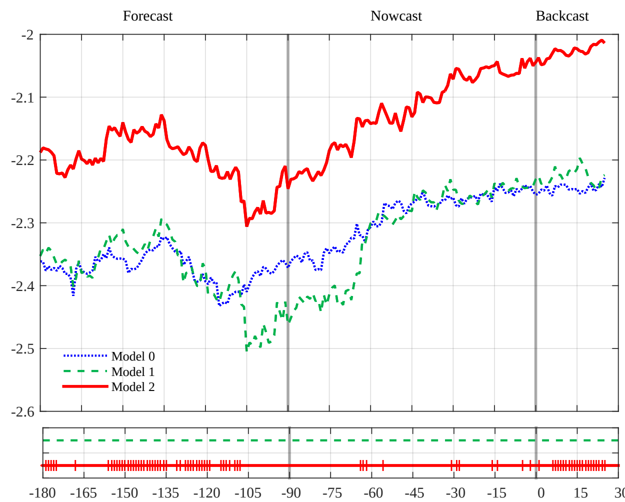
Figure H.3: Density Forecast Accuracy Evaluation

(a) Log Probability Score

Full Sample: 2000:Q1-2015:Q1

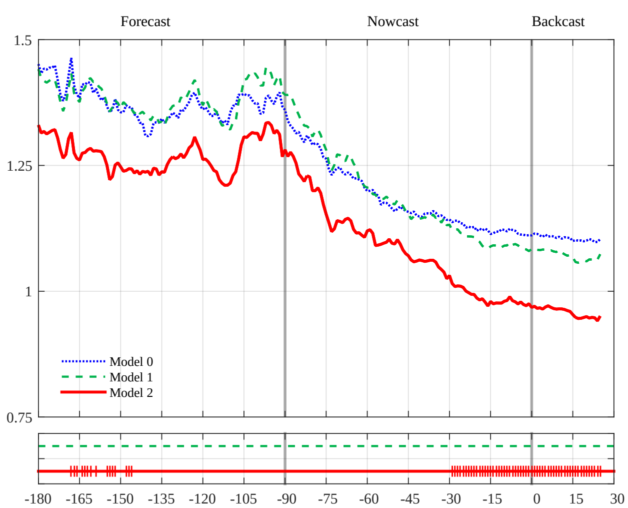


Recovery Sample: 2009:Q2-2015:Q1

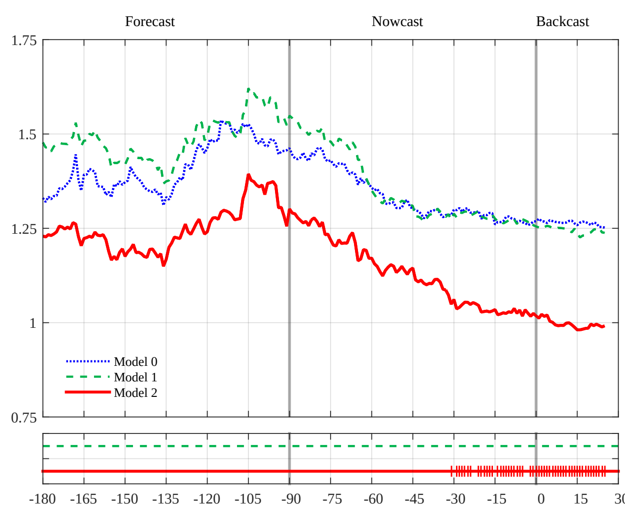


(b) Continuous Rank Probability Score

Full Sample: 2000:Q1-2015:Q1



Recovery Sample: 2009:Q2-2015:Q1

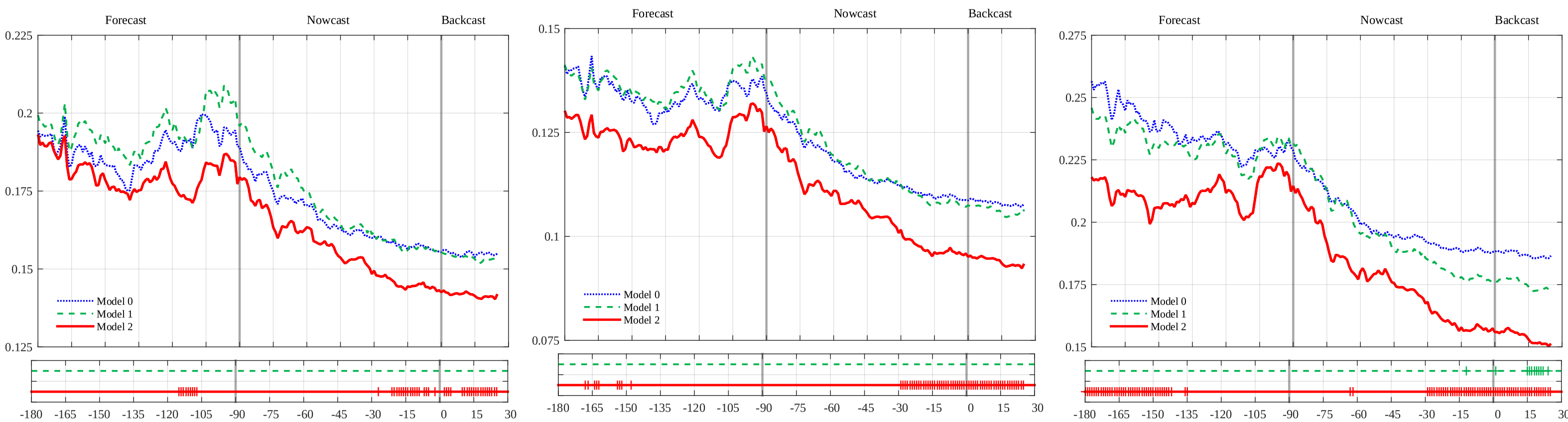


Note: The horizontal axis indicates the forecast horizon, expressed as the number of days to the end of the reference quarter. Thus, from the point of view of the forecaster, forecasts produced 180 to 90 days before the end of a given quarter are a forecast of next quarter; forecasts 90-0 days are nowcasts of current quarter, and the forecasts produced 0-25 days after the end of the quarter are backcasts of last quarter. The boxes below each panel display, with a vertical tick mark, a gauge of statistical significance at the 10% level of any difference with Model 0, for each forecast horizon, as explained in the main text.

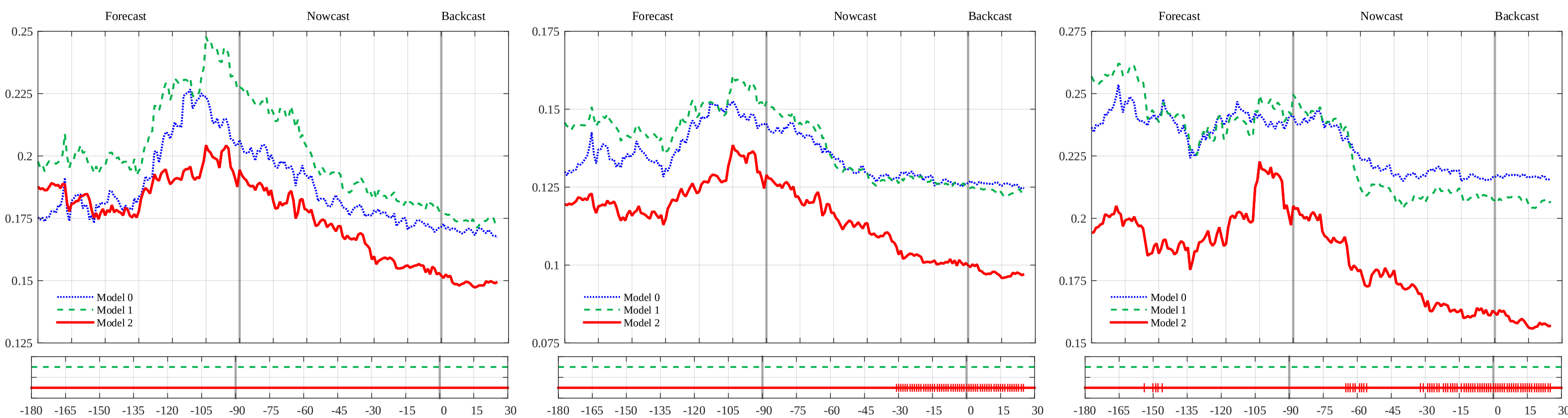
Figure H.4: Density Forecast Accuracy Evaluation: Quantile Score Statistics

Full Sample: 2000:Q1-2015:Q1

(a) Left | (b) Center | (c) Right



Recovery Sample: 2009:Q2-2015:Q1



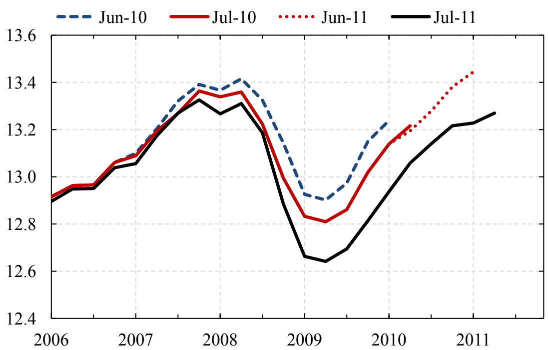
Note: The horizontal axis indicates the forecast horizon, expressed as the number of days to the end of the reference quarter. Thus, from the point of view of the forecaster, forecasts produced 180 to 90 days before the end of a given quarter are a forecast of next quarter; forecasts 90-0 days are nowcasts of current quarter, and the forecasts produced 0-25 days after the end of the quarter are backcasts of last quarter. The boxes below each panel display, with a vertical tick mark, a gauge of statistical significance at the 10% level of any difference with Model 0, for each forecast horizon, as explained in the main text.

I Case Study - The Decline of The Long-Run Growth Estimate in Mid-2010 and Mid-2011

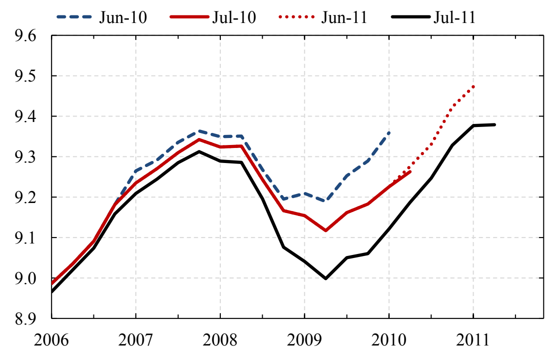
Figure I.1 looks in more detail at the specific information that, in real time, led the model to reassess its estimate of long-run growth. There are large reassessments of long-run growth around July 2010 and July 2011, coinciding with the publication by the Bureau of Economic Analysis of the annual revisions to the National Accounts, which each year incorporate previously unavailable information for the previous three years. In both cases, the revisions implied substantial downgrades both to GDP (Panel a) and in particular to the growth of consumption (Panel b) in the first years of the recovery, from 2.5% to 1.6%, and instead allocated much of the growth in GDP during the recovery to inventory accumulation. The estimate of long-run growth produced by our model is downgraded sharply as information about these revision is coming in, reflecting the role of consumption as the most persistent and forward looking component of GDP. This is clearly visible in Panels (c) and (d) of Figure I.1. In particular, Panel (c) presents the evolution of the GDP nowcast for 2010 produced by Model 2 (black line), in comparison with the counterfactual nowcast that would result if there had been no revisions to long-run growth (dashed line). It is evident that the bulk of the revisions to GDP growth that year are the consequence of a large downward revision to long-run growth. Panel (d) plots the annual nowcast of GDP produced by Model 0 (dashed line), which does not allow for changes in long-run GDP growth, and Model 2 (solid line), our baseline specification. Up to mid-2010, both models produce similar nowcasts (not shown). After 2010, however, it is clearly visible that Model 0 begins each year expecting robust growth of above 3%, only to be disappointed by incoming data. The nowcasts by Model 2, which has incorporated the decline in long-run growth, do not suffer from the same pattern of systematic downward surprises.

Figure I.1: Case Study: Impact of National Accounts Revision

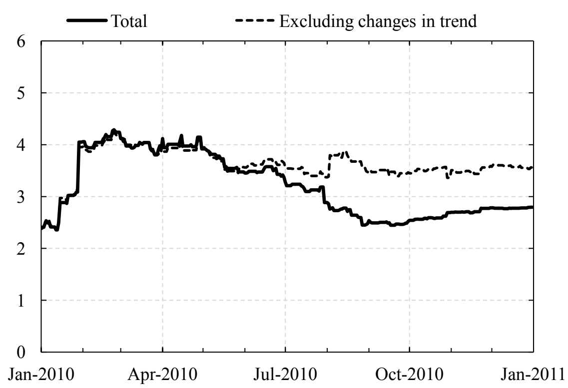
(a) Vintages of Real GDP



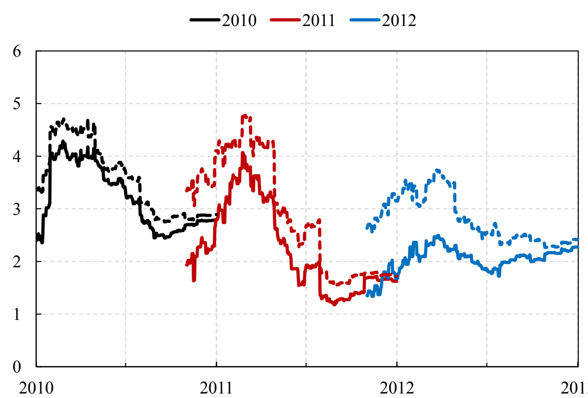
(b) Vintages of Real Personal Consumption Exp.



(c) Breakdown of Model 2's nowcast for 2010



(d) Evolution of calendar-year GDP nowcasts



Note: Panels (a) and (b) compare several vintages of data on real GDP and real personal consumption expenditures around the time of important revisions by the BEA. Panel (c) presents the evolution of the GDP nowcast for 2010 produced by Model 2 (black line), in comparison with the counterfactual nowcast that would result if there had been no revisions to long-run growth (dashed line). The evolution of calendar-year nowcasts of real GDP growth produced by Model 0 (dashed) and Model 2 (solid) are presented in Panel (d).

J Inspecting Data Set Size and Composition - More Details

J.1 Extended Model: Estimation Using a Very Large Panel

With regards to the size of the data set, in Section 4.1 of the main text we argue in favor of excluding disaggregated series within the various categories of real activity. This is because of the fact that the strong correlation across series within the same category might be a source of model misspecification. This is for two reasons: first, strong correlation in the idiosyncratic terms of series between the same category, and second, the fact that finer disaggregation levels are available for certain categories can lead to oversampling, see Boivin and Ng (2006) and Alvarez et al. (2012) for more details.

It is possible, however, to consider a more general specification of our model that can alleviate this problem, once we take into account the fact that persistent idiosyncratic movements common across series of the same category usually reflect differences in phase relative to the common activity factor. For example, all series related to employment respond to innovations to real activity with a lag. An interesting question is how our results are affected if one were to aim to make the dimension of $\mathbf{y}_t$ as large as possible, instead of carefully making variable selection based on the criteria discussed in the paper. In order to illustrate this point, we construct a "universe" of potentially available real activity time series for inclusion, based on a systematic attempt to find as many as possible US real activity time series. This is the "extended model" introduced in Section 4.6 of the paper.

J.1.1 Construction of the Extended Data Set

To construct the data panel for the extended model, we proceed as follows. First, we obtain all of the monthly real activity variables contained in the data set used by Stock and Watson (2012), which results in 75 time series.[10] Second, we exhaustively expand the monthly series contained in our original data set along all levels and dimensions of disaggregation available through Haver Analytics.[11] Out of this collection of expanded series, we then select any series that is not already contained in the 75 Stock and Watson indicators. Overall, this procedure results in a data set of as many as 155 time series capturing US real activity.[12]

J.1.2 Extended Model Specification

Maintaining the specification with a single factor (i.e. $k = 1$) we modify equation (1) of the paper as follows:

$$\mathbf{y}_t = \mathbf{c}_t + \mathbf{\Lambda}(L) f_t + \mathbf{u}_t, \tag{15}$$

[10]Details on this data set can be found in the online supplement to Stock and Watson (2012), available on Mark Watson's website. The only variable we were not able to obtain is Construction Contracts, which is not publicly available.

[11]This includes for example disaggregation along sectoral, regional and demographic characteristics.

[12]A detailed list of variables is available upon request.

such that the loading matrix $\Lambda(L)$ is now a polynomial in the lag operator of order s, i.e. contains the loadings on the contemporaneous common factor and its lags. In the special case where $s = 0$ we obtain our baseline specification. For the extended model, we set the maximum lag length, $s = 5$. The remaining equations of the model remain unchanged.

J.1.3 Priors and Model Settings

The data is standardized prior to estimation. “Minnesota”-style priors are applied to the coefficients in $\mathbf{\Lambda}(L)$, $\phi(L)$ and $\rho_i(L)$. More specifically:

- For the autoregressive coefficients of the factor dynamics, $\phi(L)$, the prior mean is set to 0.9 for the first lag, and to zero in subsequent lags. This reflects a belief that the common factor captures a highly persistent but stationary business cycle process.

- For the factor loadings, $\mathbf{\Lambda}(L)$, the prior mean is set to 1 for the first lag, and to zero in subsequent lags. This shrinks the model towards the factor being the cross sectional average of the variables, see D’Agostino et al. (2015).

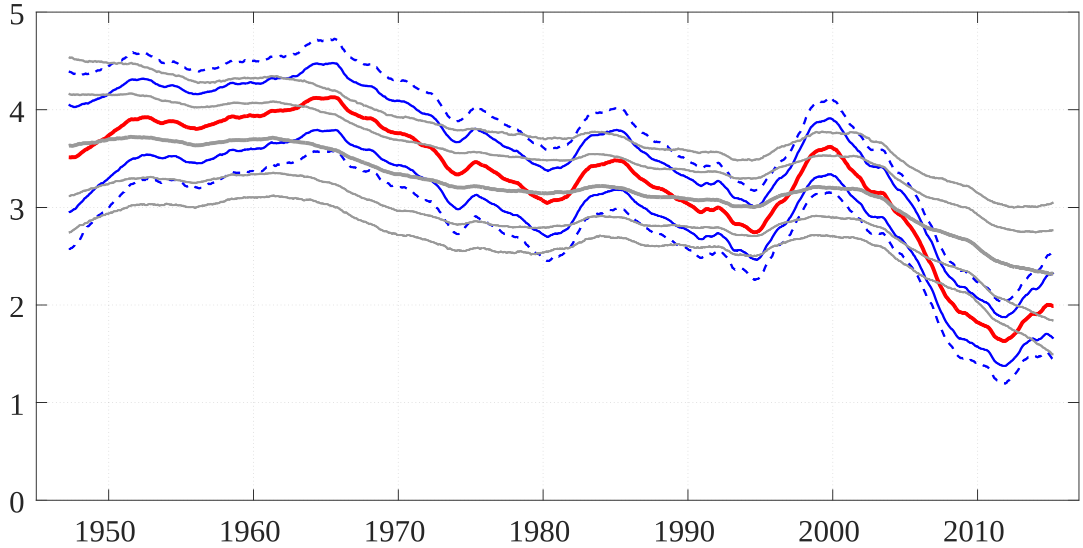
- For the autoregressive coefficients of the idiosyncratic, $\rho_i(L)$ the prior is set to zero for all lags, thus shrinking the model towards a model with no serial correlation in $u_{i,t}$.

In all cases, the variance on the priors is set to $\frac{\tau}{h^2}$, where τ is a parameter governing the tightness of the prior, and h is equal to the lag number of each coefficent, ranging $1 : p$, $1 : q$ and $1 : s + 1$. Following D’Agostino et al. (2015), we set $\tau = 0.2$, a value which is standard in the Bayesian VAR literature.

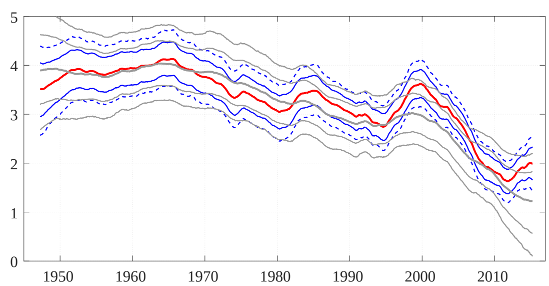
J.2 Results Across Alternative Data Sets

Figure J.1: Comparison Across Alternative Data Sets/Models

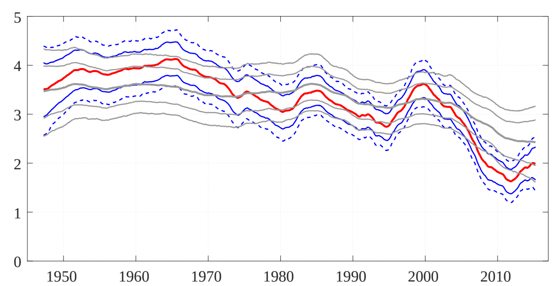
(a) Baseline With And Without Including Consumption



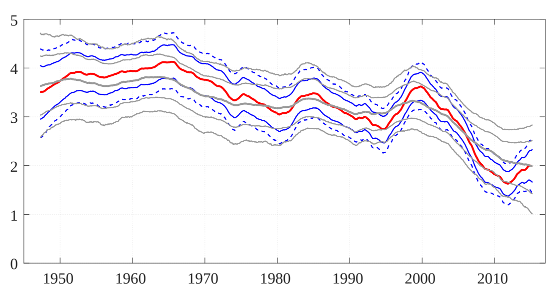
(b) Okun



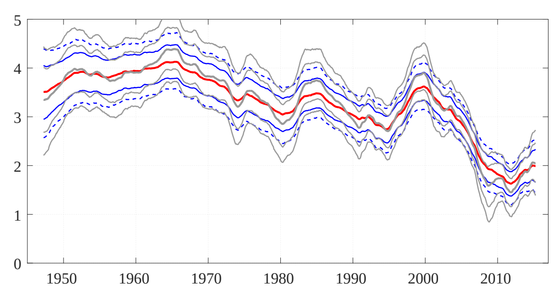
(c) Mariano-Murasawa



(d) Extended model ex. consumption



(e) Extended model



Note: In each panel our baseline the median estimate of real GDP growth is presented (red), with corresponding 68% (solid blue) and the 90% (dashed blue) posterior credible intervals. The corresponding estimates for the respective alternative data sets are superimposed in gray.

K A Growth Accounting Exercise

The decomposition exercise carried out in Section 5 of the paper provides a first step towards giving an economically more meaningful interpretation of the movements in long-run real GDP growth detected by our model. While our equation $g_t = z_t + h_t$ follows from a simple identity, we demonstrate in this appendix how it can be related to the standard growth accounting framework.

To illustrate this point, consider two versions of the standard neoclassical growth model. In the first version, assume a standard Cobb-Douglas production function with Hicks-neutral technological change and constant returns to scale. In growth rates, this can be written as

$$dlogY_t = dlogTFP_t + \alpha dlogK_t + (1-\alpha)dlogH_t, \tag{16}$$

where Y_t, K_t and H_t denote the level of output, the capital stock and labor input (total hours), respectively. α is the capital share and TFP_t is total factor productivity. Rearranging this relation gives

$$dlogY_t = dlogTFP_t + dlogH_t + \alpha(dlogK_t - dlogH_t), \tag{17}$$

so that the growth rate of real GDP is the sum of long-run growth in technology, total hours and a third term which captures differential growth in input factors which implies changes in the capital-labor ratio ("capital deepening"). In the second version of the neoclassical growth model, consider adding growth in labor-augmenting technology in the form of labor quality, denoted Q_t. In this case, the relation between growth rates is rearranged to

$$dlogY_t = dlogTFP_t + dlogH_t + \alpha(dlogK_t - dlogH_t) + (1-\alpha)dlogQ_t. \tag{18}$$

Both relations (17) and (18) can be captured in our econometric framework. We define the first four elements of our vector of observables $\mathbf{y}_t$ in equation (1) to be the growth rate in real GDP, real consumption, TFP and total hours worked. As in the baseline model, *transitory* fluctuations in inputs (due to temporary shocks) would still be captured by the cyclical dactor, f_t, whereas the various sources of *permanent* changes in the growth rate of inputs (say, the long-run growth rate of technology, or the long-run growth rate of the population) would be included in $\mathbf{a}_t$. In order to mimic the relations prescribed by the two versions of the neoclassical growth model, we specify the long-run time variation in our model, $\mathbf{a}_t$ as a composite of three terms. While $\tilde{h}_t$ captures long-run movements in hours, the movements in long-run labor productivity are now further decomposed into a "technology" term $\tilde{z}_t$ and a "non-technology" term $\tilde{x}_t$. Formally, $\mathbf{c}_t$ in equation (2) is constructed as

$$\mathbf{a}_t = \begin{bmatrix} \tilde{z}_t \\ \tilde{h}_t \\ \tilde{x}_t \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{19}$$

What the non-technology term corresponds to depends on the underlying structure that is assumed. For instance, in the first version of the neoclassical growth model above

$$\tilde{x}_t \equiv \alpha(dlogK_t - dlogH_t) \tag{20}$$

and in the second case

$$\tilde{x}_t \equiv \alpha(dlogK_t - dlogH_t) + (1-\alpha)dlogQ_t. \tag{21}$$

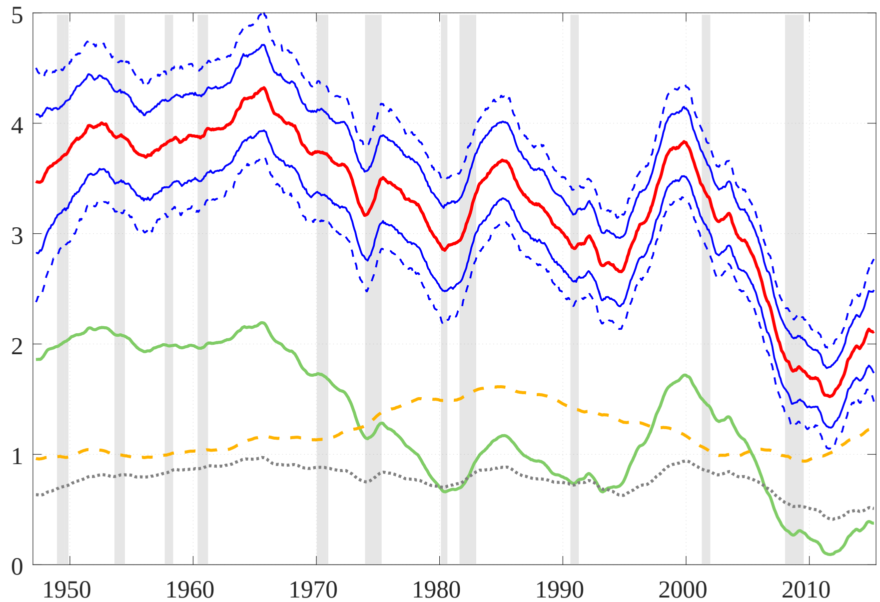
In both cases, $\tilde{x}_t$ subsumes potential long-run factors other than TFP that may explain changes in the long-run labor productivity trend we discuss in the paper.[13] $\tilde{z}_t$ is intended to capture changes in the long-run technology growth rate.

Figure K.1 presents the results for US data when defining $\mathbf{c}_t$ by (19), and the measure of utilization-adjusted TFP from Fernald (2012) is used as an additional observable. Panel (a) shows the posterior estimate of long-run real GDP growth (including bands), together with the decomposition into long-run total hours growth, long-run technology growth and long-run non-technological growth. Reassuringly, the evolution of the total long-run growth component, g_t (red) is virtually identical to the estimate from our baseline model. The estimate of long-run hours growth (orange) is also very similar to its counterpart in Section 5 of the paper. Interestingly, the non-technological term (dashed gray) is positive on average and is relatively stable over the sample. Finally, the key insight from panel (a) is that our estimate of the long-run technology (green) displays strong movements that are very similar to the long-run growth rate of labor productivity which we have extracted in the simpler decomposition in Section 5. Under the assumption of a neoclassical structure, changes in long-run technology growth appear to be the main driver behind the recent slowdown in long-run real GDP growth. Panel (b) plots the growth rate of the utilization-adjusted TFP measure by Fernald (2012) in black, together with its long-run counterpart as estimated by our model (green, with blue bands). It is visible that the DFM approach is capable of extracting a smooth low frequency trend from the volatile series of TFP, which captures well-known episodes such as the 1970's slowdown and the IT boom of the 1990's. Overall, our framework is capable of providing an interesting angle on real-time movements in technology trends.

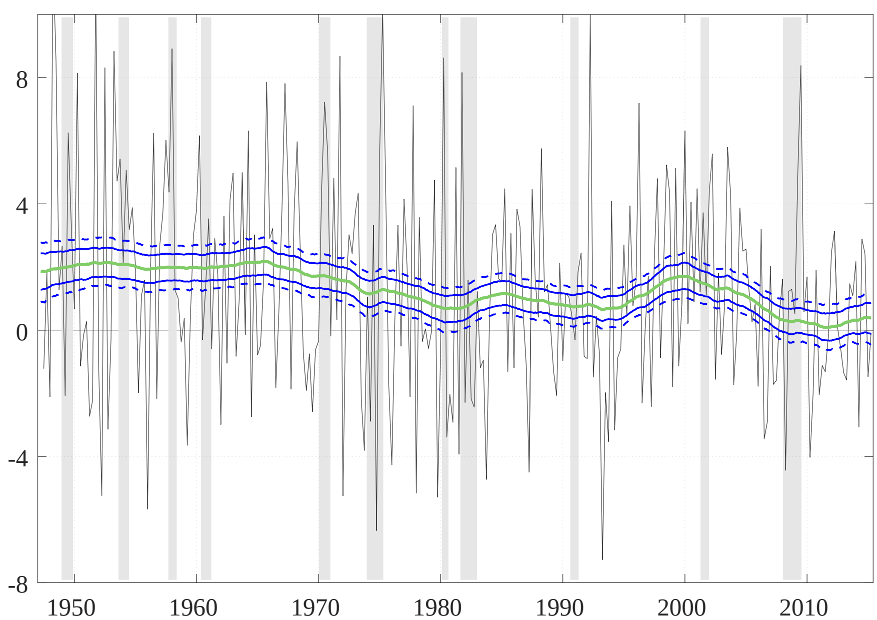
[13]Note here that in the first case we could also directly capture the technological parameter α into the matrix $\mathbf{B}$ by setting its (1,4) and (2,4) elements to α and interpreting $\tilde{z}_t$ directly as capital deepening. The specification above is somewhat more appealing in that it allows for a non-constant capital share. One can easily impose a constant value for α by scaling the posterior estimate of $\tilde{x}_t$ by that value.

Figure K.1: Results of Growth Accounting Exercise

(a) Further Decomposition of US Long-Run US Output Growth



(b) Fernald (2012) TFP Growth: Data and Long-Run Estimate



Note: Panel (a) displays the posterior median estimates of long-run real GDP growth in red, together with the posterior median estimates of its components, long-run hours growth, long-run TFP growth and long-run non-technological growth (orange dashed, green, gray dotted). For long-run real GDP growth the corresponding with corresponding 68% and 90% posterior credible intervals are shown in solid and dashed blue. Panel (b) plots the growth rate of utilization-adjusted TFP by Fernald (2012) in black, together with its long-run counterpart in our econometric framework, i.e. the estimate of $\tilde{z}_t$, with corresponding 68% and 90% posterior credible intervals (green/blue).

References

Alvarez, R., Camacho, M., and Perez-Quiros, G. (2012). Finite sample performance of small versus large scale dynamic factor models. CEPR Discussion Papers 8867, C.E.P.R. Discussion Papers.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–68.

Bai, J. and Wang, P. (2015). Identification and Bayesian Estimation of Dynamic Factor Models. *Journal of Business & Economic Statistics*, 33(2):221–240.

Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2012). Now-casting and the real-time data flow. Working Papers ECARES 2012-026, ULB.

Banbura, M. and Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160.

Beveridge, S. and Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics*, 7(2):151–174.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Carter, C. K. and Kohn, R. (1994). On Gibbs Sampling for State Space Models. *Biometrika*, 81(3):541–553.

CBO (2001). CBOs Method for Estimating Potential Output: An Update. Working papers, The Congress of the United States, Congressional Budget Office.

Clark, T. E. and McCracken, M. W. (2012). In-sample tests of predictive ability: A new approach. *Journal of Econometrics*, 170(1):1–14.

Cogley, T. and Sargent, T. J. (2005). Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S. *Review of Economic Dynamics*, 8(2):262–302.

D'Agostino, A., Giannone, D., Lenza, M., and Modugno, M. (2015). Nowcasting business cycles: A bayesian approach to dynamic heterogeneous factor models. Technical report.

de Wind, J. and Gambetti, L. (2014). Reduced-rank time-varying vector autoregressions. CPB Discussion Paper 270, CPB Netherlands Bureau for Economic Policy Analysis.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4):863–83.

Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.

Faust, J. and Wright, J. H. (2013). Forecasting inflation. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2.

Fernald, J. (2012). A quarterly, utilization-adjusted series on total factor productivity. *Federal reserve bank of San Francisco working paper*, 19:2012.

Giacomini, R. and White, H. (2006). Tests of Conditional Predictive Ability. *Econometrica*, 74(6):1545–1578.

Gneiting, T. and Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102:359–378.

Gneiting, T. and Ranjan, R. (2011). Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules. *Journal of Business & Economic Statistics*, 29(3):411–422.

Groen, J. J. J., Paap, R., and Ravazzolo, F. (2013). Real-Time Inflation Forecasting in a Changing World. *Journal of Business & Economic Statistics*, 31(1):29–44.

Hansen, B. E. (1992). Testing for parameter instability in linear models. *Journal of Policy Modeling*, 14(4):517–533.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Jacquier, E., Polson, N. G., and Rossi, P. E. (2002). Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics*, 20(1):69–87.

Kim, C.-J. and Nelson, C. R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. The MIT Press.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65(3):361–93.

Marcellino, M., Porqueddu, M., and Venditti, F. (2014). Short-term gdp forecasting with a mixed frequency dynamic factor model with stochastic volatility. *Journal of Business & Economic Statistics, Forthcoming*.

Mariano, R. S. and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18(4):427–443.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84(405):223–230.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852.

Stock, J. and Watson, M. (2012). Disentangling the channels of the 2007-09 recession. *Brookings Papers on Economic Activity*, Spring:81–156.