Following the Trend: Tracking GDP when Long-Run Growth is Uncertain*

Juan Antolin-Diaz
Fulcrum Asset Management

Thomas Drechsel
LSE, CFM

Ivan Petrella
Bank of England, Birkbeck, CEPR

This draft: January 28, 2015
First draft: October 24, 2014

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 this slowdown started 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, variations in the growth rate of labor productivity appear to be the most important driver of changes 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, Marble Arch House, 66 Seymour Street, London W1H 5BT, UK; E-Mail: juan.antolin-diaz@fulcrumasset.com. Drechsel: Department of Economics and Centre for Macroeconomics, London School of Economics, Houghton Street, London, WC2A 2AE, UK; E-Mail: t.a.drechsel@lse.ac.uk. Petrella: Structural Economic Analysis Division, Bank of England, Threadneedle Street, London EC2R 8AH, UK; E-Mail: Ivan.Petrella@bankofengland.co.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 or the Bank of England, or to state Bank of England policy. This paper should therefore not be reported as representing the views of the Bank of England or members of the Monetary Policy Committee or Financial Policy Committee. We thank Neele Balke, Gavyn Davies, Wouter den Haan, Ed Hoyle, Juan Rubio-Ramirez, Ron Smith, Paolo Surico and the seminar participants at ECARES and BBVA Research 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 persistently produced negative forecast errors over the last five 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.[2] 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 large number of economic activity indicators, DFMs are capable of precisely estimating the cylical comovements in macroeconomic data in a real-time setting. Our model exploits this to track changes in the long-run growth rate of GDP in a timely and reliable manner, separating them from their cyclical counterpart.

The evidence of a decline in long-run US GDP growth is accumulating, as documented by the recent growth literature such as Fernald and Jones (2014). Lawrence

[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]See for example Auerbach (2011).

Summers and Robert Gordon have been articulating a rather pessimistic view of long-run growth which contrasts with the optimism prevailing before the Great Recession (see Jorgenson et al., 2006). To complement this evidence, we start the analysis by presenting the results of two popular structural break tests, Nyblom (1989) and Bai and Perron (1998). Both suggest that a possible shift in the mean of US 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-2014, 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 GDP growth 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 stemming from 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 2.25%, almost one per-

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

centage point lower than the postwar average. The results are more consistent with 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 vintages of data. By the summer of 2011 the model would have concluded that a significant decline in long-run growth was behind the slow recovery, well before the structural break tests became conclusive. Since the seminal contributions of Evans (2005) and Giannone et al. (2008) DFMs have become the standard tool to track GDP.[4] Taking into account the variation in long-run GDP growth improves substantially the point and density GDP "nowcasts" produced by this class of models. Furthermore, we show that our DFM provides an advantage over traditional trend-cycle decompositions in detecting changes in the long-run growth rate of GDP by using a larger amount of information.

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, by adding information about aggregate hours worked, long-run output growth can be decomposed into labor productivity and labor input trends.[5] 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

[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] An alternative approach to extract the long run component of productivity would be to extract this directly from measured labour productivity (see e.g. Roberts, 2001, Benati, 2007, and Edge et al., 2007). Our approach has the advantage that hours are clearly more cyclical and less noisy than measured productivity, which naturally enhances the precision of the estimated long run components.

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.[6] Acknowledging the importance of allowing for time-variation in the means of the variables, Stock and Watson (2012) pre-filter their dataset 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 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 obtain a substantially less pessimistic and more precise estimate of the long-run growth of GDP than these studies in the latest part of the sample, which we attribute to the larger amount of information we incorporate on cyclical developments.

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.

[6]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.

2 Preliminary Evidence

The literature on economic growth reveals 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 therefore long-run output growth occurred in the early 1970's (for a retrospective see Nordhaus, 2004), and that faster productivity in the IT sector led to an acceleration in the late 1990's (Oliner and Sichel, 2000). 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 econometric models. After the slowdown in productivity became apparent in the 1970's, many papers 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.

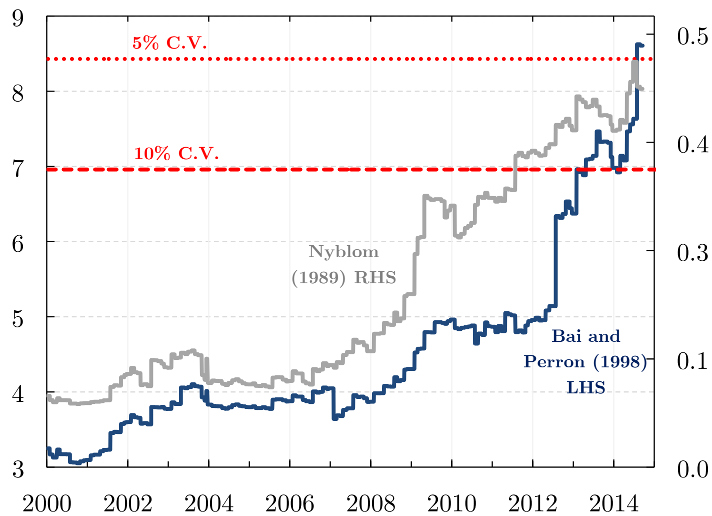
During the recovery from the Great Recession US GDP has grown well below its postwar average. There are two popular strategies that could be followed from a frequentist perspective to detect parameter instability or the presence of breaks in the mean growth rate. The first one is Nyblom's (1989) L-test 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 1960-2014 this test rejects the stability of the constant term at 10% whereas the stability of the autoregressive coefficient cannot be rejected.[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 mean of real US GDP growth at the 5%-level. The most likely break is in the second quarter of 2000.[8] Using the latter methodology, 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 highlight that substantial evidence for a recent change in the mean of US GDP growth since the 1960's 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. This reveals that a break would not have been detected at the 5% significance levels until as late as mid-2014, which is almost fifteen years later than the actual break data suggested by the Bai and Perron (1998) procedure. The Nyblom (1989) test, which is designed to detect gradual change, first becomes significant at the 10%-level in 2011, somewhat faster than the discrete break

[7] The same results hold for an AR(2) specification. In both cases, the stability of the variance is rejected at the 1%-level. Interestingly, McConnell and Perez-Quiros (2000) also use this test and show that in their sample there is evidence for instability only in the variance but not in the constant term or the autoregressive coefficient of the AR(1).

[8] The second most likely break, which is not statistically significant, is estimated to have occurred in the second quarter of 1973. Appendix A provides the full results of the tests and further discussion.

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



Note: The solid gray and blue 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 1960 and the test is re-applied for every new data release occurring after the beginning of 2000. The dotted line plots the 5% critical value of the test, while the dashed line plots the 10% critical value.

test. This highlights the importance of an econometric framework capable of quickly adapting to changes in long-run growth as new information arrives.

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 very large datasets, 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, one objective of this paper is to make these models robust to changes in long-run growth. Moreover, we propose that the argument that information contained in a broad panel of monthly indicators improves the precision of short-run GDP forecasts extends to the estimation of the time-varying long-run growth rate of GDP. In essence, as the number of indicators becomes large, the cyclical component of GDP growth is estimated arbitrarily well, facilitating the decomposition of residual variations into persistent movements in long-run growth and short-lived noise. Section 4.5 expands in more detail on the reasoning behind this key argument for choosing the DFM framework.

While we remain agnostic about the ultimate form of structural change in the GDP process, we propose specifying its long-run growth rate as a random walk. Our motivation is similar to Primiceri (2005). While in principle it is unrealistic to conceive that GDP growth 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 the trend as a random walk is more robust to misspecification when the actual process is indeed characterized by discrete breaks, whereas models with discrete breaks might not be robust to the true process being a random walk.[9] Finally, the random walk assumption also has the desirable feature that, unlike stationary models, confidence bands around GDP forecasts increase with the forecast horizon, reflecting uncertainty about the possibility of future breaks or drifts in long-run growth.

[9] We demonstrate this point with the use of Monte Carlo simulations, showing that a random walk trend 'learns' quickly about a large 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. See Appendix B for the full results of these simulations.

3.1 The Model

Let y_t be an $(n \times 1)$ vector of observable macroeconomic time series, and let f_t denote a $(k \times 1)$ vector of latent common factors. It is assumed that $n >> k$, so that the number of observables is much larger than the number of factors. Setting $k = 1$ and ordering GDP growth first (therefore GDP growth is referred to as $y_{1,t}$) we have[10]

$$y_{1,t} = \alpha_{1,t} + f_t + u_{1,t}, \tag{1}$$

$$y_{i,t} = \alpha_i + \lambda_i f_t + u_{i,t}, \quad i = 2, \ldots, n \tag{2}$$

where $u_{i,t}$ is an idiosyncratic component specific to the i^{th} series and λ_i is its loading on the common factor.[11] Since the intercept $\alpha_{1,t}$ is time-dependent in equation (1), we allow the mean growth rate of GDP to vary. We choose to do so only for GDP, which is sufficient to track changes in its long-run growth while keeping the model as parsimonious as possible.[12] If some other variable in the panel was at the center of the analysis or there was suspicion of changes in its mean, an extension to include additional time-varying intercepts would be straightforward. In fact, for theoretical reasons it might be desirable to impose that the drift in long-run GDP growth is shared by other series such as consumption, a possibility that we consider in Section 5,

[10]For the purpose of tracking GDP with a large number of closely related indicators, the use of one factor is appropriate (see Section 4.2) and we therefore focus on the case of $k = 1$ for expositional clarity in this section.

[11]The loading for GDP is normalized to unity. This serves as an identifying restriction in our estimation algorithm. Bai and Wang (2012) discuss minimal identifying assumptions for DFMs.

[12]The alternative approach of including a time-varying intercept for all indicators (see, e.g. Creal et al., 2010 or Fleischman and Roberts, 2011) implies that the number of state variables increases with the number of observables. This not only imposes an increasing computational burden, but in our view compromises the parsimonious structure of the DFM framework, in which the number of degrees of freedom does not decrease as more variables are added. It is also possible that allowing for time-variation in a large number of coefficients would improve in-sample fit at the cost of a loss of efficiency in out-of-sample forecasting. For the same reason we do not allow for time-variation in the autoregressive dynamics of factors and idiosyncratic components, given the limited evidence on changes in the duration of business cycles (see e.g. Ahmed et al., 2004).

where we also include a time-varying intercept for aggregate hours worked and explore the underlying drivers of the long-run growth decline.

The laws of motion for the factor and idiosyncratic components are, respectively,

$$\Phi(L)f_t = \varepsilon_t, \tag{3}$$

$$\rho_i(L)u_{i,t} = \eta_{i,t} \tag{4}$$

$\Phi(L)$ and $\rho_i(L)$ denote polynomials in the lag operator of order p and q, respectively. Both (3) and (4) are covariance stationary processes. The disturbances are distributed as $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2_{\varepsilon,t})$ and $\eta_{i,t} \overset{iid}{\sim} N(0, \sigma^2_{\eta_{i,t}})$, where the SV is captured by the time-variation in $\sigma^2_{\varepsilon,t}$ and $\sigma^2_{\eta_{i,t}}$.[13] The idiosyncratic components, $\eta_{i,t}$, are cross-sectionally orthogonal and are assumed to be uncorrelated with the common factor at all leads and lags. Finally, the dynamics of the model's time-varying parameters are specified to follow driftless random walks:

$$\alpha_{1,t} = \alpha_{1,t-1} + v_{\alpha,t}, \qquad v_{\alpha,t} \overset{iid}{\sim} N(0, \varsigma_{\alpha,1}) \tag{5}$$

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

$$\log \sigma_{\eta_{i,t}} = \log \sigma_{\eta_{i,t-1}} + v_{\eta_{i,t}}, \qquad v_{\eta_{i,t}} \overset{iid}{\sim} N(0, \varsigma_{\eta,i}) \tag{7}$$

where $\varsigma_{\alpha,1}$, ς_{ε} and $\varsigma_{\eta,i}$ are scalars.[14]

Note that in the standard DFM, it is assumed that ε_t and $\eta_{i,t}$ are iid. Moreover,

[13] Once SV is included in the factors, it must be included in all idiosyncratic components as well. In fact, the Kalman filter estimates of the state vector will depend on the signal-to-noise ratios, $\sigma_{\varepsilon_t}/\sigma_{\eta_{i,t}}$. If the numerator is allowed to drift over time but the denominator is kept constant, we might be introducing into the model spurious time-variation in the signal-to-noise ratios, implying changes in the precision with which the idiosyncratic components can be distinguished from the common factors.

[14] For the case of more than one factor, following Primiceri (2005), the covariance matrix of f_t, denoted by $\Sigma_{\varepsilon,t}$, can be factorized without loss of generality as $A_t \Sigma_{\varepsilon,t} A'_t = \Omega_t \Omega'_t$, where A_t is a lower triangular matrix with ones in the diagonal and covariances $a_{ij,t}$ in the lower off-diagonal elements, and Ω_t is a diagonal matrix of standard deviations $\sigma_{\varepsilon_{i,t}}$. Furthermore, for $k > 1$, ς_{ε} would be an unrestricted $(k \times k)$ matrix.

both the factor VAR in equation (3) and the idiosyncratic components (4) are usually assumed to be stationary, so by implication the elements of y_t are assumed to be stationary (i.e. the original data have been differenced appropriately to achieve stationarity). In equations (1)-(7) we have relaxed these assumptions to allow for a stochastic trend in the mean of GDP and SV. Our model nests specifications that have been previously proposed for tracking GDP. We obtain the DFM with SV of Marcellino et al. (2014) if we shut down time variation on the mean of GDP, i.e. set $\varsigma_{\alpha,1} = 0$. If we further shut down the SV, i.e. set $\varsigma_{\alpha,1} = \varsigma_{\epsilon} = \varsigma_{\eta,i} = 0$, we obtain the specification of Banbura and Modugno (2014) and Banbura et al. (2012).

3.2 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 the growth rate of GDP, which is quarterly, and 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 rate can be related to the (unobserved) monthly growth rate and its lags using a weighted mean:

$$y^q_{1,t} = \frac{1}{3}y^m_{1,t} + \frac{2}{3}y^m_{1,t-1} + y^m_{1,t-2} + \frac{2}{3}y^m_{1,t-3} + \frac{1}{3}y^m_{1,t-4}, \tag{8}$$

where only every third observation of $y^q_{1,t}$ is actually observed. Substituting (1) into (8) 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. Below we will present a Bayesian estimation method that 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).

3.3 State Space Representation and Estimation

The model features autocorrelated idiosyncratic components (see equation (4)). In order to cast it in state-space form, 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 2012).[15] Specifically, defining $\bar{y}_{i,t} \equiv (1 - \rho_i(L))y_{i,t}$ for $i = 2, \ldots, n$ and $\tilde{y}_t = [y_{1,t}, \bar{y}_{2,t}, \ldots, \bar{y}_{n,t}]'$, the model can be compactly written in the following state-space representation:

$$\tilde{y}_t = HX_t + \tilde{\eta}_t, \tag{9}$$

$$X_t = FX_{t-1} + e_t, \tag{10}$$

where the state vector stacks together the time-varying intercept, the factors, and the idiosyncratic component of GDP, as well as their lags required by equation (8). To be precise, $X_t = [\alpha_{1,t}, \ldots, \alpha_{1,t-4}, f_t, \ldots, f_{t-mp}, u_{1,t}, \ldots, u_{1,t-mq}]'$, where $mp = \max(p, 4)$ and $mq = \max(q, 4)$. Therefore the measurement errors, $\tilde{\eta}_t' = [0, \bar{\eta}_t']$ with $\bar{\eta}_t = [\eta_{2,t}, \ldots, \eta_{n,t}]' \sim N(0, R_t)$, and the transition errors, $e_t \sim N(0, Q_t)$, are not serially correlated. The system matrices H, F, R_t and Q_t depend on the hyperparameters of the

[15] As an alternative, Banbura and Modugno (2014) suggest including these components as additional elements of the state vector. This solution has the undesirable feature that the number of state variables will increase with the number of observables, leading to a loss of computational efficiency.

DFM, $\lambda, \Phi, \rho, \sigma_{\varepsilon,t}, \sigma_{\eta_{i,t}}, \varsigma_{\alpha_1}, \varsigma_{\varepsilon}, \varsigma_{\eta}$.

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 (2012), but extend it to include mixed frequencies, the time-varying intercept, and SV.[16] The SVs are sampled using the approximation of Kim et al. (1998), which is considerably faster than the alternative Metropolis-Hasting algorithm of Jacquier et al. (2002). Our complete sampling algorithm together with the details of the state space representation can be found in Appendix C.

4 Evidence for US Data

4.1 Priors and Model Settings

We wish to impose as little prior information as possible. In our baseline results 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 $\varsigma_{\alpha,1}$, ς_{ε} and $\varsigma_{\eta,i}$ in equations (5)-(7) are however difficult to identify from the information contained in the likelihood function alone. As the literature on Bayesian VARs documents, attempts to use non-informative priors for these parameters will in many cases produce relatively high posterior estimates, i.e. 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 benchmark

[16] Simulation algorithms in which the Kalman Filter is used over thousands of replications frequently produce a singular covariance matrix due to the accumulation of rounding errors. Bai and Wang (2012) propose a modification of the well-known Carter and Kohn (1994) algorithm to prevent this problem which improves computational efficiency and numerical robustness. We thank Jushan Bai and Peng Wang for providing the Matlab code for the square-root form Kalman Filter.

model, which excludes time-varying long-run GDP growth and SV. In particular, we set an inverse gamma prior with one degree of freedom and scale equal to 0.001 for $\varsigma_{\alpha,1}$.[17] For ς_ϵ and ς_η we set an inverse gamma prior with one degree of freedom and scale equal to 0.0001.

In our empirical application the number of lags in the polynomials $\Phi(L)$ and $\rho(L)$ is set to 2 ($p = 2$ and $q = 2$ respectively) in the spirit of Stock and Watson (1989). The model can be easily extended to include more lags in both transition and measurement equations, i.e. to allow the factors to load some variables with a lag. In the latter case, it is again sensible to avoid overfitting by choosing priors that shrink the additional lag coefficients towards zero (see e.g. D'Agostino et al., 2012, and Luciani and Ricci, 2014).

4.2 Data

A number of studies on DFMs, including Giannone et al. (2005), Banbura et al. (2012), Alvarez et al. (2012) and Banbura and Modugno (2014) highlight that the inclusion of nominal or financial variables, of disaggregated series beyond the main headline indicators, or the use of more than one factor do not meaningfully improve the precision of real GDP forecasts. We follow them in focusing on a medium-sized panel of real activity data including only series for each economic category at the highest level of aggregation, and set the number of factors $k = 1$. The single factor can in this case be interpreted as a coincident indicator of economic activity (see e.g. Stock and Watson, 1989, and Mariano and Murasawa, 2003). Relative to the latter studies, which include just four and five indicators respectively, the conclusion of the literature is that adding additional indicators, in particular surveys, does improve the

[17] To gain an intuition about this prior, note that over a period of ten years, this would imply that the posterior mean 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.

precision of short-run GDP forecasts (Banbura et al., 2010). A key finding of our paper is that survey indicators are also informative to separate the cyclical component of GDP growth from its long-run counterpart. This is because in many cases these surveys are by construction stationary, and have a high signal-to-noise ratio, which provides a clean signal of the cycle excluding movements in long-run growth.

Our panel of 26 data series is shown in Table 1. Since we are interested in covering a long sample in order to study the fluctuations in long-run growth, we start our panel in January 1960. Here 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 which start as late as 2007.[18]

4.3 In-Sample Results

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.[19] Panel (a) of Figure 2 plots the posterior median, together with the 68% and 90% posterior credible intervals of the long-run growth rate. This estimate is conditional on the entire sample and accounts for both filtering and parameter uncertainty. For comparison, the well-known estimate of potential growth produced by the Congressional Budget Office (CBO) is also plotted. Several features of our estimate of long-run growth are worth noting. An initial slowdown is visible around the late

[18]Our criteria for data selection is similar to the one proposed by Banbura et al. (2012), who suggest including the headline series that are followed closely by financial market participants. In practice, we consider that a variable is widely followed by markets when survey forecasts of economists are available on Bloomberg prior to the release. Some surveys appear to be better aligned with the rest of the variables after taking a 12 month difference transformation, a feature that is consistent with these indicators sometimes being regarded as leading rather than coincident.

[19]Thanks to the efficient state space representation discussed above, the improvements in the simulation smoother proposed by Bai and Wang (2012), 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 a standard Intel 3.6 GHz computer with 16GB of DDR3 Ram.

Table 1:
Data series used in empirical analysis

	Freq.	Start Date	Transformation	Publ. Lag
Hard Indicators				
Real GDP	Q	Q1:1960	% QoQ Ann.	26
Industrial Production	M	Jan 60	% MoM	15
New Orders of Capital Goods	M	Mar 68	% MoM	25
Light Weight Vehicle Sales	M	Feb 67	% MoM	1
Real Personal Consumption Exp.	M	Jan 60	% MoM	27
Real Personal Income less Trans. Paym.	M	Jan 60	% MoM	27
Real Retail Sales Food Services	M	Jan 60	% MoM	15
Real Exports of Goods	M	Feb 68	% MoM	35
Real Imports of Goods	M	Feb 69	% MoM	35
Building Permits	M	Feb 60	% MoM	19
Housing Starts	M	Jan 60	% MoM	26
New Home Sales	M	Feb 63	% MoM	26
Payroll Empl. (Establishment Survey)	M	Jan 60	% MoM	5
Civilian Empl. (Household Survey)	M	Jan 60	% MoM	5
Unemployed	M	Jan 60	% MoM	5
Initial Claims for Unempl. Insurance	M	Jan 60	% MoM	4
Soft Indicators				
Markit Manufacturing PMI	M	May 07	-	-7
ISM Manufacturing PMI	M	Jan 60	-	1
ISM Non-manufacturing PMI	M	Jul 97	-	3
Conf. Board: Consumer Confidence	M	Feb 68	Diff 12 M.	-5
U. of Michigan: Consumer Sentiment	M	May 60	Diff 12 M.	-15
Richmond Fed Mfg Survey	M	Nov 93	-	-5
Philadelphia Fed Business Outlook	M	May 68	-	0
Chicago PMI	M	Feb 67	-	0
NFIB: Small Business Optimism Index	M	Oct 75	Diff 12 M.	15
Empire State Manufacturing Survey	M	Jul 01	-	-15

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 $y_t - y_{t-12}$. 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.

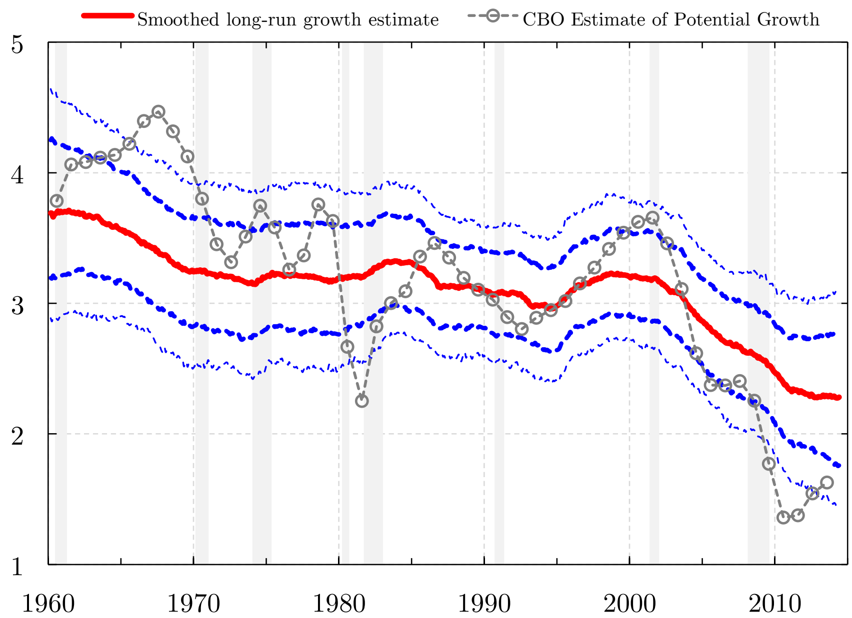
1960's, close to the 1973 "productivity slowdown" (Nordhaus, 2004). The acceleration of the late 1990's and early 2000's associated with the productivity boom in the IT sector (Oliner and Sichel, 2000) is also clearly visible. Thus, until the middle of the decade of the 2000's, our estimate conforms well to the generally accepted narrative about fluctuations in potential growth. It must be noted, however, that according to our estimates until the most recent part of the sample, the historical average of 3.15% is always contained within the 90% credible interval. Finally, from its peak of about 3.25% in late 1998 to its level as of June 2014, 2.25%, the median estimate of the trend has declined by one percentage point, a more substantial decline than the one observed after the original productivity slowdown of the 1970's. Moreover, the slowdown appears to have happened gruadually since the start of the 2000's, with about half of the total decline having occurred before the Great Recession and the rest immediately after.

Our estimate of long-run growth and the CBO's capture similar 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 Beverdige 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.[20] It is nevertheless interesting that despite employing different statistical methods they produce qualitatively similar results, 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, about half of the slowdown occurred prior to the Great Recession. The CBO's estimate was

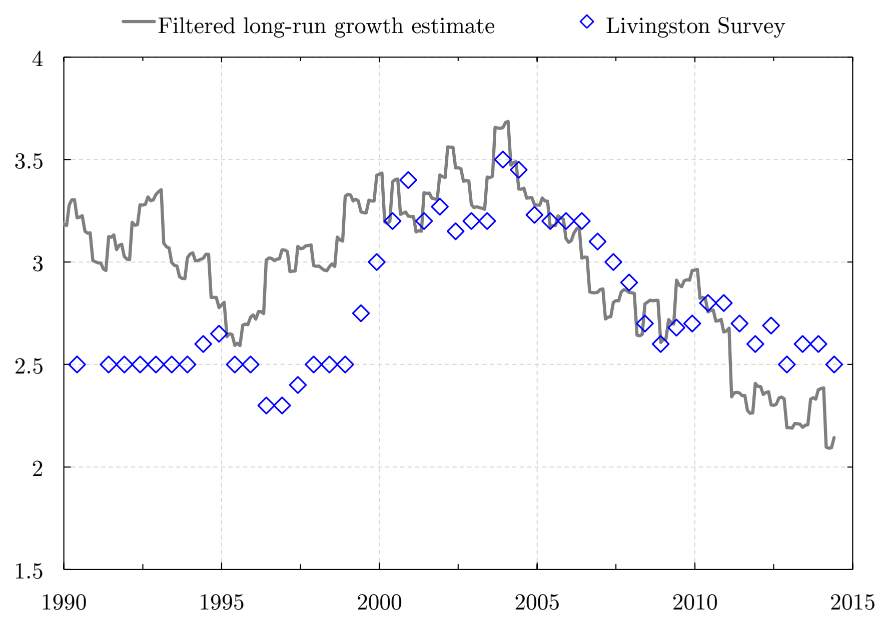
[20]Essentially, the production function approach calculates the trend components of the supply inputs to a neoclassical production function (the capital stock, total factor productivity, and the total amount of hours) using statistical filters and then aggregates them to obtain an estimate of the trend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

Figure 2: US long-run growth estimate: 1960-2014 (% Annualized Growth Rate)

(a) Posterior long-run growth estimate vs CBO estimate of potential growth



(b) Filtered estimates of long-run growth vs SPF survey



Note: Panel (a) plots the posterior median (solid red), together with the 68% and 90% (dashed blue) posterior credible intervals of long-run GDP growth. The gray circles are the CBO's estimate of potential growth. Shaded areas represent NBER recessions. In Panel (b), the solid gray line is the filtered estimate of the long-run GDP growth rate, $\hat{\alpha}_{1,t|t}$, using the vintage of National Accounts available as of mid-2014. The blue 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.

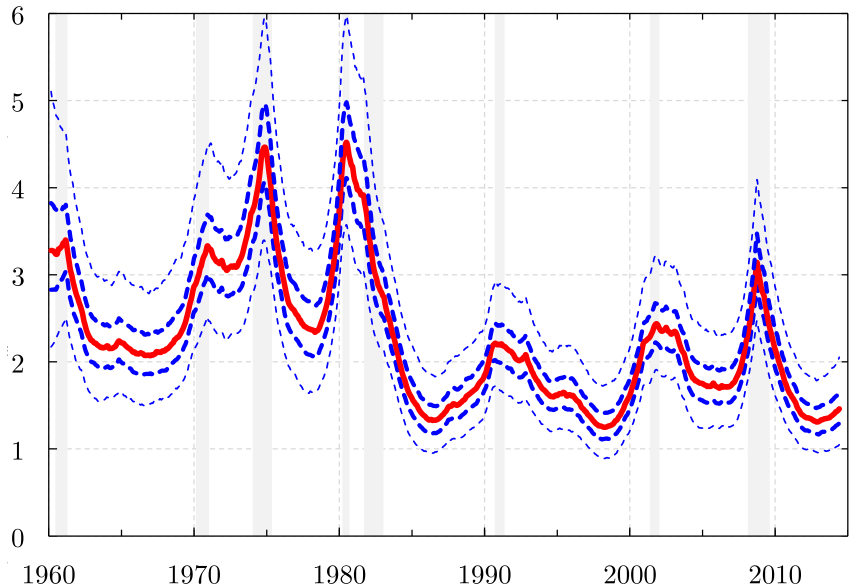
significantly below ours immediately after the recession, reaching an unprecedented low level of about 1.25% in 2010, and remains in the lower bound of our posterior estimate since then. Section 4.5 expands on the reason for this divergence and argues that it stems from the larger amount of information incorporated in the DFM.

The posterior estimates, $\hat{\alpha}_{1,t|T}$, are outputs of a Kalman smoother recursion, i.e. they are conditioned on the entire sample, so 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 Appendix B, we show that if changes in long-run growth occur in the form of discrete breaks rather than evolving gradually, additional insights can be obtained looking at the filtered estimates, $\hat{\alpha}_{1,t|t}$, which will tend to jump after a break. Panel (b) shows that the filtered estimate of long-run growth is still consistent with a relatively gradual slowdown. The model's estimate declines from about 3.5% in the early 2000's to about 2.25% as of mid-2014. The largest downgrades occur in 2003 and in the summer of 2011. As an additional external benchmark we also include the real-time median forecast of average real GDP growth over the next ten years from the Livingston Survey of Professional Forecasters' (SPF). It is noticeable that the SPF was substantially more pessimistic during the 1990's, and did not incorporate the substantial acceleration in trend growth due to the 'New Economy' until the end of the decade. From 2005 to about 2010, the two estimates are remarkably similar, showing a deceleration to about 2.75% as the productivity gains of the IT boom faded. This matches the narrative of Fernald (2014). Since then, the SPF forecast has remained relatively stable whereas our model's estimate has declined by a further half percentage point.

Figure 3 presents the posterior estimate of the SV of the common factor.[21] The

[21] To be precise, this is the square root of $\text{var}(f_t) = \sigma^2_{\varepsilon,t}(1-\phi_2)/[(1+\phi_2)((1-\phi_2)^2-\phi_1^2)]$.

Figure 3: Stochastic Volatility of Common Factor



Note: The figure presents the median (red), the 68% (solid blue) and the 90% (dashed blue) posterior credible intervals of the idiosyncratic component of the common factor. Shaded areas represent NBER recessions.

Great Moderation is clearly visible, with the average volatility pre-1985 being about twice 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 seems to spike during recessions, a finding that brings our estimates close to the recent findings of Jurado et al. (2014) and Bloom (2014) relating to business-cycle uncertainty.[22] It appears that the random walk specification is flexible enough to capture cyclical changes in volatility as well as permanent phenomena such as the

[22]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.

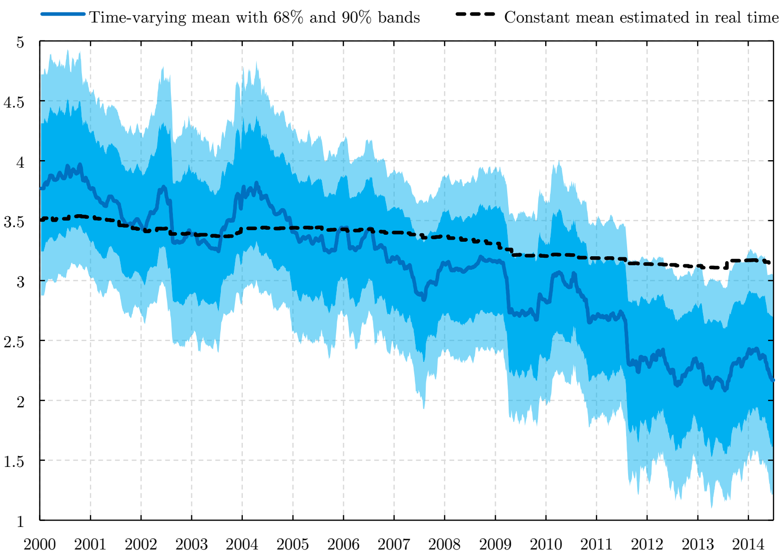
Great Moderation. Appendix D provides analogous charts for the estimated volatilities of the idiosyncratic components of selected data series. Similar to the volatility of the common factor, many of the idiosyncratic volatilities present sharp increases during recessions.

The above results provide evidence that a significant decline in long-run US GDP growth occurred over the last decade, and are consistent with a relatively gradual decline since the early 2000's. Both smoothed and filtered estimates show that around half 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. The other half took place in the aftermath of the recession. The overall decline is the largest and most significant movement in long-run growth observed in the postwar period.

4.4 Implications for Tracking GDP in Real Time

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 possible that our results are only apparent using the latest vintage of data, and our model would not have been able to detect the slowdown in long-run growth as it happened. To address this concern, we reconstruct our dataset at each point in time, using vintages of data available from the Federal Reserve Bank of St. Louis ALFRED database. 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 1960:Q1 and use an expanding out-of-sample window which starts on 11 January 2000 and ends in 22 September 2014. 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.[23]

Figure 4: Long-Run GDP Growth Estimates in Real Time



Note: The shaded areas represent the 68th and 90th percentile, together with the median of the posterior credible interval of the current value of long-run GDP growth, re-estimated daily using the vintage of data available at each point in time from January 2000 to September 2014. The dashed line is the contemporaneous estimate of the historical average of GDP growth rate. In both cases the start of the sample is fixed at Q1:1960.

Figure 4 presents the model's real-time assessment of the posterior distribution of long-run growth at each point in time. The estimate of long-run growth from a model with a constant intercept for GDP growth is plotted for comparison. This estimate is also updated as new information arrives, but weighs all points of the sample equally. The evolution of the model's estimate of long-run growth when estimated in real time is remarkably similar to the in-sample results discussed above. About half of the

[23] 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 sequentially apply the selection criterion of Banbura et al. (2012) and append to the panel, in real time, the vintages of the new indicators as soon as Bloomberg surveys of forecasters are available. In the example of the PMI, surveys appear in Bloomberg 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 Appendix E.

total decline was detected already by December 2007, and by the summer of 2011 a large enough decline has occurred such that almost the entire probability mass of the posterior distribution is below the historical average.

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 our specification with time-varying long-run growth and SV also improves the performance of the model along this dimension. We find that over the evaluation window 2000-2014 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 2014. Appendix F provides the full details of the forecast evaluation exercise.

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.

4.5 The Role of Information in Trend-Cycle Decompositions

It is interesting to asses whether using a DFM that includes information from a broad panel of monthly indicators provides an advantage over traditional trend-cycle decompositions in detecting changes in the long-run growth rate of GDP. Most studies usually focus on a few cyclical indicators, generally inflation or variables that are themselves inputs to the production function (see e.g. Gordon, 2014a or Reifschneider et al., 2013), whereas DFMs can benefit from incorporating a potentially large number of indicators and weighing them optimally according to their signal-to-noise ratio. In this paper we argue that as the number of indicators becomes large, the cyclical component of GDP growth is estimated arbitrarily well, facilitating the trend-cycle decomposition through the inclusion of more information.[24]

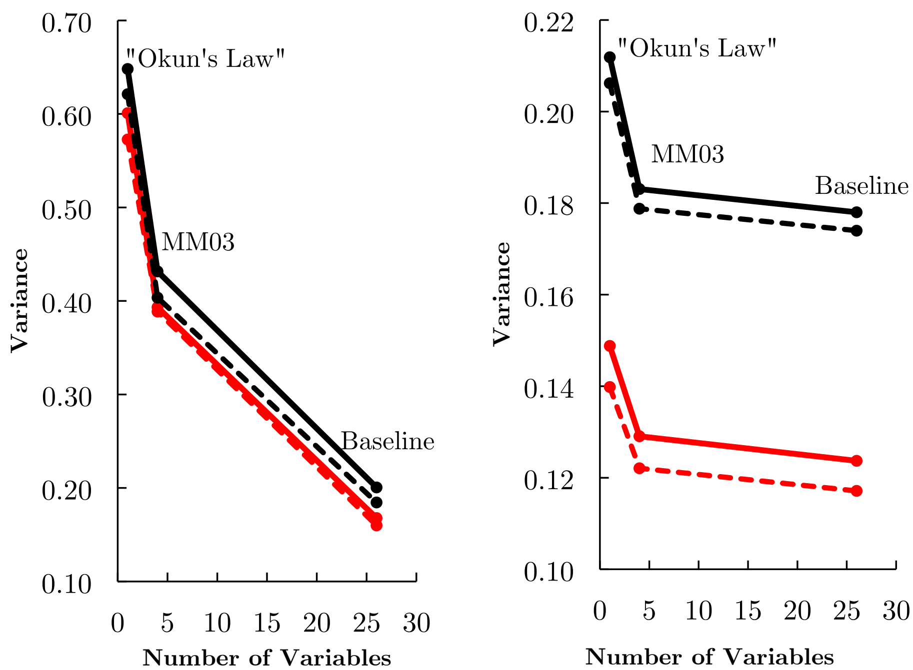
Figure 5 compares the uncertainty around the estimates of the common factor and the long-run GDP growth component from our model and alternative DFMs differing only in number of cyclical indicators included. In particular, we consider (1) a model with GDP and unemployment only (labeled "Okun's Law"), (2) an intermediate model with GDP and the four additional variables included in Mariano and Murasawa (2003) and Stock and Watson (1989) (labeled "MM03"), and (3) our baseline specification. For each model we compute the average variance of the filtered and smoothed estimates, and decompose it into filtering uncertainty and parameter uncertainty following Hamilton (1986). The uncertainty around the common cyclical factor, displayed in panel (a), is dramatically reduced as more indicators are added. As visible in panel (b), part of this reduction spills over to the long-run growth component. The improvement comes from both types of uncertainty, but mostly from fitering uncertainty. The "MM03" specification already achieves a substantial reduction in uncertainty, despite using only

[24]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).

five variables. It should be noted, however, that these indicators are highly informative about cyclical developments, and yet are usually not included in studies of long-run growth based on the production function. The model put forward in our paper reduces the uncertainty somewhat further and has the additional advantage of including variables, such as surveys, that are available in a more timely manner, leading to a reduction of uncertainty in real time.

Figure 5: Uncertainty in Different Trend-Cycle Decompositions

(a) Common Cyclical Factor (b) Long-Run Growth Component



Note: This figure compares the uncertainty surrounding the composition into trend and cycle in three alternative models, where the horizontal axis depicts the number of variables used. Panel (a) displays the uncertainty around the cyclical component, whereas the uncertainty around the long-run growth component is visible in panel (b). In both panels, the black lines correspond to the smoothed estimate and the red lines to the filtered estimates. The solid line captures the total uncertainty coming from both filtering and parameter uncertainty. The dashed corresponds to uncertainty excluding parameter uncertainty.

The point made here helps in explaining the recent divergence between our estimate of long-run GDP growth and the CBO's potential growth estimate. The CBO's method embeds an Okun's Law relationship, by which growth is above potential when the unemployment rate is declining. The combination of weak GDP growth and a strong decline in unemployment of the last few years has therefore led to drastic downward revisions of their potential GDP growth. However, many observers have noted that the strong decline in labor force participation invalidates the use of the unemployment rate as a sufficient indicator of cyclical conditions (see e.g. Erceg and Levin, 2013). By extracting the cyclical factor from a broad panel of activity variables, our model is more robust to an abnormal development in one particular variable.

4.6 Robustness to Priors and Model Settings

Recall that in our main results we use non-informative priors for loadings and serial correlation coefficients, and a conservative prior for the time variation in the long-run growth of GDP and SV. The non-informative priors were motivated by our desire to make our estimates more comparable to the existing literature. As a robustness check we also consider a "Minnesota"-style prior on the autoregressive coefficients of the factor,[25] as well as shrinking the coefficients of the serial correlation towards 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.[26] As for the prior on the amount variation of long-run growth, our choice of a conservative prior was motivated by our desire to minimize overfitting, especially given that the GDP series, being observed only at the quarterly frequency, has one third of

[25] To be precise, we center the prior on the first lag around 0.9 and the subsequent lags at zero.

[26] There is some evidence that the use of these 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.

the observations of the monthly variables. Other researchers might however believe that the permanent component of GDP growth is more variable, so we re-estimated the model using a larger prior for the variance of its time variation (i.e. $\varsigma_{\alpha,1} = 0.01$). As a consequence, the long-run component displays a more pronounced cyclical variation but it is estimated less accurately. Interestingly, the increase in the variance and cyclicality of the long-run growth component brings it closer to the CBO estimate.

5 Decomposing Movements in Long-Run Growth

So far we have argued that in order to track movements in long-run GDP growth, and to ensure that GDP forecasts are robust to changes in the latter, it is sufficient to include a time-varying intercept in the GDP equation only. In this section, we show that with minimal modifications, 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 the underlying long-run trends. We see this exercise as a tentative step towards giving an economic interpretation to the movements in long-run 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. By adding the hours series to the data panel, it is 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.[27] In the measurement equation of the DFM, the first of these components loads output but not hours, and the second loads

[27]Throughout this section, we construct quarterly series for aggregate hours worked in the total economy following the methodology of Ohanian and Raffo (2012). In particular, we benchmark the series of average hours provided by the BLS to the annual estimates compiled by the Conference Board's Total Economy Database (TED), which are regarded as more reliable and comparable across countries.

both series. We interpret the first component as the long-run growth rate of labor productivity, while the second one captures low-frequency movements in labor input independent of productivity. The state-space representation presented in Section 3.3 allows us to implement this decomposition in the form of restrictions on the loading matrix.[28] Formally, let $y_{1,t}$ and $y_{2,t}$ denote the growth rate of real GDP and total hours worked, respectively. Equation (1) and the first element of equation (2) are then re-written as

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z_t \\ h_t \end{bmatrix} + \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix} f_t + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}, \tag{11}$$

where z_t and h_t jointly follow random walks with diagonal covariance matrix. The measurement equation for the other monthly indicators $y_{i,t}$ for $i = 3, \ldots, n$ remains unchanged. Given equation (11), the two components add up to the time-varying intercept in the baseline specification, i.e. $\alpha_{1,t} = z_t + h_t$. Moreover, growth in output per hour, $y_{1,t} - y_{2,t}$, is equal to z_t in the long run.[29] Since z_t and h_t are independent and add up to $\alpha_{1,t}$, we set the prior on the scale of their variances to half of the one set in Section 4.1 on $\alpha_{1,t}$.

The above restrictions identify z_t and h_t, but in order to reduce uncertainty around their estimates, it is desirable to add additional information on these components that might be contained in consumption, an argument stemming from a large strand of the literature. Motivated by the permanent income hypothesis, Harvey and Stock (1988), Cochrane (1994) and Cogley (2005) argue that incorporating information about consumption is informative about the permanent component in GDP.[30] Since consumption

[28] The restrictions are implemented by drawing from the restricted conditional posterior. See Bai and Wang (2012) for details.

[29] In addition, note that the cyclical movement in labor productivity is given by $(1 - \lambda_2)f_t$.

[30] Importantly, consumption of durable goods should be excluded given that the ratio of their price index to the GDP deflator exhibits a downward trend, and therefore the chained quantity index grows faster than overall GDP. Following Whelan (2003), for this section we construct a Fisher index of non-durables and services and use its growth rate as an observable variable in the panel.

and output should grow at the same rate in the long-run, we impose that z_t and h_t also load consumption growth with coefficient equal to unity.

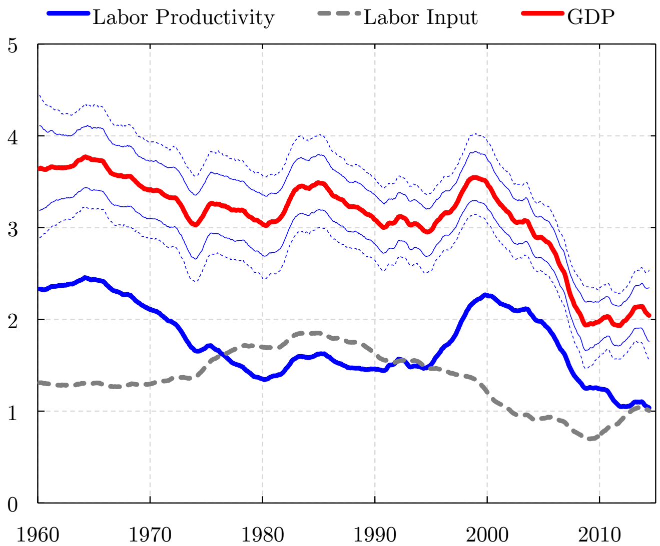
Figure 6 presents the results of the decomposition exercise for the US. Panel (a) plots the median posterior estimate of long-run GDP growth and its labor productivity and total hours components. The posterior bands for long-run GDP growth are included.[31] The time series evolution conforms very closely 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. This contrasts with the popular narrative 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.[32] Panel (b) plots 2,000 draws from the joint posterior distribution of the variances of the innovations to the labor productivity and hours components. About 70% of the draws lie above the 45° line, while under the equal-variance prior

[31] Note that including consumption reduces the uncertainty around the estimate of long-run GDP growth relative to Figure 2 and makes its movements somewhat more pronounced. In particular, the entire slowdown in growth now appears to have occurred before the Great Recession. A comparison is available in Appendix D.

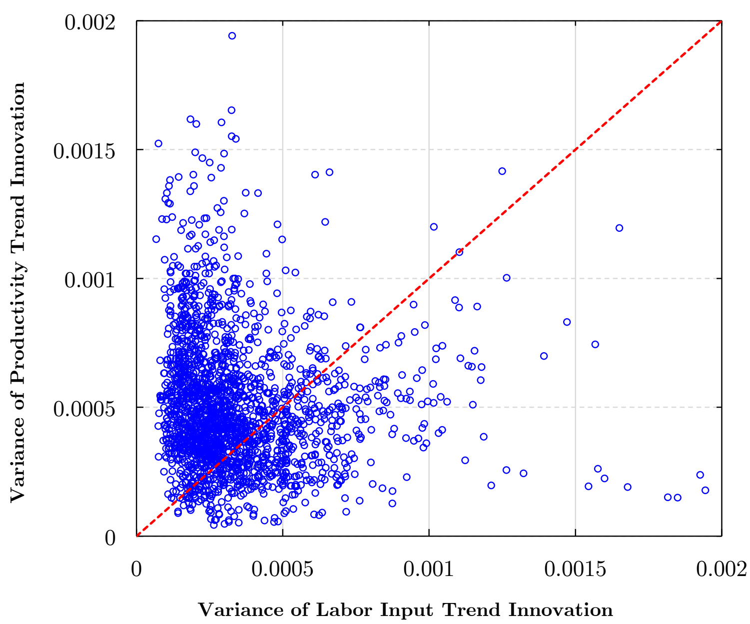
[32] The filtered estimates of the two components, available in Appendix D, reveal that around the turn of the century the slowdown in GDP growth is explained by a gradual decline in the labor input component. Starting around 2005 a more abrupt revision to labor productivity drives the decline in overall long-run GDP. A comparison with the SPF shows how the difference in the overall growth rate displayed in Figure 2 is explained by a larger downward revision of long-run labor productivity. The end of the sample is associated with a rise in the labor input component accompanied by stagnating labor productivity.

Figure 6: Decomposition of Long-run US Output Growth

(a) Posterior median estimates of decomposition



(b) Joint posterior distribution of component's innovation variances



Note: Panel (a) plots the posterior median (red), together with the 68% and 90% (dashed blue) 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 (solid blue and dashed grey lines). Panel (b) plots 2,000 draws of the joint posterior distribution of the variances of innovations to the labor productivity and hours components. The dashed red line is the 45°-line.

the draws would be equally distributed above and below it. This confirms the visual impression from Panel (a) that changes in labor productivity, rather than in labor input, are the key driver of low frequency movements in real GDP growth.

It is interesting to compare the results of our decomposition exercise to similar approaches in the recent 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. The key difference is that we use a large amount of information, allowing 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 4.5. Another 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 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. In terms of results, 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%. Since Gordon's model can be nested into ours, we attribute this difference to our use of a large-dimensional system.[33]

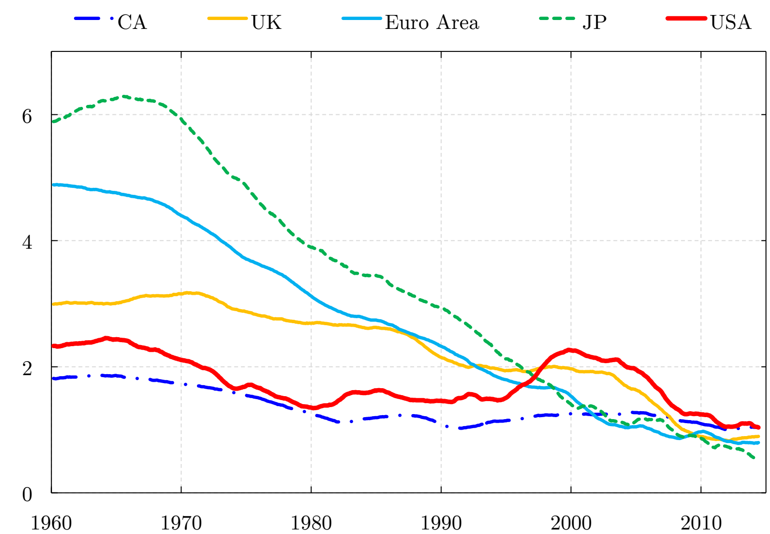
To gain an international perspective on our results, we estimate the DFM for the other G7 economies and perform the decomposition exercise for each of them.[34] The

[33]In Appendix D we provide the results for a bivariate model of GDP and unemployment and show that the current long-run growth estimate is 1.3%, close to Gordon (2014a).

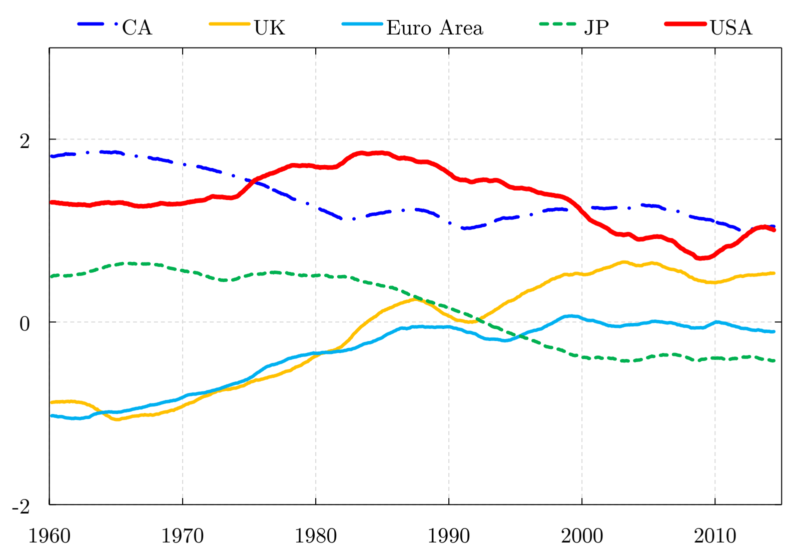
[34]Details on the specific data series used for each country are available in Appendix E. For hours, we again follow the methodology of Ohanian and Raffo (2012). In the particular case of the UK, the quarterly series for hours displays a drastic change in its stochastic properties in the early 1990's 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 (8) to annual variables (see Banbura et al. 2012). To avoid weak

Figure 7: 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.

median posteriors of the labor productivity and labor input trends are displayed in Figure 7. Labor productivity, displayed in Panel (a), plays again the key role in determining movements in long-run. 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 were bound to end 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. Interestingly, this acceleration in the 1990's did not occur in Western Europe and Japan.[35] The UK stands out for the unprecedented decline in labor productivity around the time of the Great Recession. 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 (11). 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. Whatever these might be, 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 according to

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.

[35]See also Hayashi and Prescott (2002) and Gordon (2004).

our model entirely accounted for by a decline in the productivity trend. Finally, it is interesting to note 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.[36]

6 Concluding Remarks

The sluggish recovery from the Great Recession has raised the question whether the long-run growth rate of US 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 2.25% in 2014. 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 2011 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 GDP 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 slowdown in labor productivity as the main driver of weak global growth in recent years, extending the narrative of Fernald (2014) to other economies. Our econometric approach remains agnostic about the deep structure of the economy.

[36] 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.

Therefore a fully specified structural model would be required to go beyond the simple decomposition presented here and to attach a more precise interpretation to the low-frequency movements in macroeconomic aggregates. However, we have shown that these long-run movements are an important feature of the data that can be successfully modeled within the DFM framework without compromising its appealing parsimonious structure.

References

Ahmed, S., Levin, A., and Wilson, B. A. (2004). Recent U.S. Macroeconomic Stability: Good Policies, Good Practices, or Good Luck? *The Review of Economics and Statistics*, 86(3):824–832.

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. (2012). Identification and estimation of dynamic factor models. *Working Paper*.

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

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.

Beverdige, 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.

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

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

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., and Lenza, M. (2012). The bayesian dynamic factor model. Technical report, ULB – Universite Libre de Bruxelles.

Del Negro, M. and Otrok, C. (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles. Staff Reports 326, Federal Reserve Bank of New York.

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.

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.

Erceg, C. and Levin, A. (2013). Labor Force Participation and Monetary Policy in the Wake of the Great Recession. CEPR Discussion Papers 9668, C.E.P.R. Discussion Papers.

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

Faust, J. and Wright, J. H. (2013). Forecasting inflation. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 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, Board of Governors of the Federal Reserve System (U.S.).

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.

Gadea-Rivas, M. D., Gomez-Loscos, A., and Perez-Quiros, G. (2014). The Two Greatest. Great Recession vs. Great Moderation. Working Paper 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.

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

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.

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.

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. (2014a). A new method of estimating potential real gdp growth: Implications for the labor market and the debt/gdp ratio. Working Paper 20423, National Bureau of Economic Research.

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

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.

Hamilton, J. D. (1986). State-space models. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, chapter 50, pages 3039–3080. Elsevier.

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.

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

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*, volume 1 of *MIT Press Books*. 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.

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 Federal Reserve System.

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.

Sinai, A. (2010). The business cycle in changing economy: Conceptualization, measurement, dating. *The American Economic Review*, 100(2):pp. 25–29.

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. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97:1167–1179.

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.

Appendix to "Following the Trend: Tracking GDP when Long-Run Growth is Uncertain"

A Full Results of Structural Break Tests

A.1 Bai and Perron Test

Table A.1 reports the result for the Bai and Perron (1998) test applied to US real GDP growth. 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 is rejected in each of the three tests, at the 5% level for the case of one break and at the 10% level for two and three breaks. However, we cannot reject the null of only one break against two breaks, or the null of only two against three breaks. The final test confirms the conclusion that there is evidence in favor of at least one break. The conclusions are almost identical when we use our baseline sample starting in 1960:Q1, or a longer one starting in 1947:Q1. The most likely break is identified to have happened at the beginning of 2000.

Table A.1:
Test Results of Bai-Perron Test

	1960-2014	**1947-2014**
	$SupF_T(k)$	
$k=1$	8.626**	8.582**
	[2000:Q2]	[2000:Q1]
$k=2$	5.925*	4.294
	[1973:Q2; 2000:Q2]	[1968:Q1; 2000:Q1]
$k=3$	4.513*	4.407*
	[1973:Q1; 1984:Q1; 2000:Q2]	[1968:Q4; 1982:Q3; 2000:Q1]
	$SupF_T(k\|k-1)$	
$k=2$	2.565	1.109
$k=3$	0.430	2.398
$U_d max$	8.626**	8.582**

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.

A.2 Nyblom Test

Table A.2 reports the result for the Nyblom (1989) test applied to US real GDP growth, as described in Hansen (1992). 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 A.2:
TEST RESULTS OF NYBLOM TEST

	L_c	
	AR(1)	AR(2)
μ	0.439*	0.371*
ρ_1	0.071	0.110
ρ_2		0.144
σ^2	1.102***	1.019***
Joint L_c	1.585***	1.420***

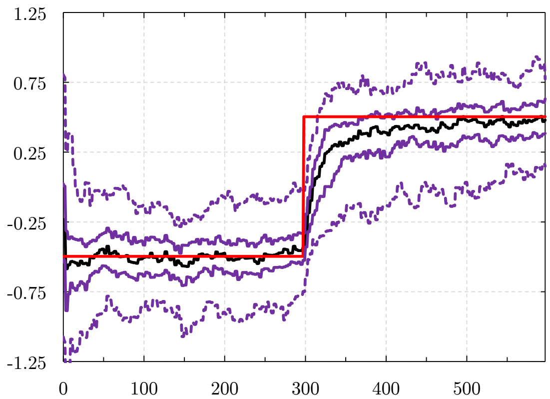
Notes: Results are obtained using Nyblom's L test as described in Hansen (1992).

B Simulation Results

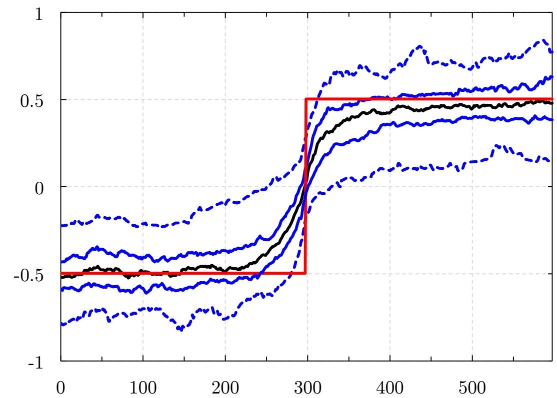
Figure B.1: Simulation Results I

Data-generating process (DGP) with one discrete break in the trend

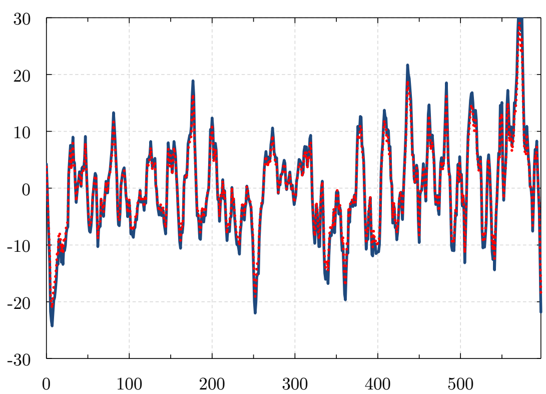
(a) True vs. Estimated Trend (Filtered)



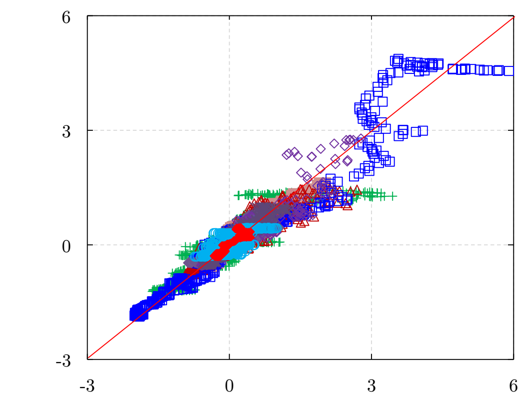
(b) True vs. Estimated Trend (Smoothed)



(c) True vs. Estimated Factor



(d) True vs. Estimated Volatilities of u_t

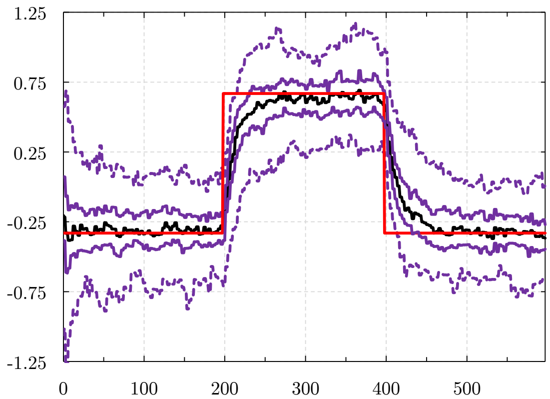


Note: The DGP features a discrete break in the trend of GDP growth occurring in the middle of the sample, as well as stochastic volatility. The sample size is $n = 26$ and $T = 600$, which mimics our US data set. The estimation procedure is the fully specified model as defined by equations (1)-(7) in the text. We carry out a Monte Carlo simulation with 100 draws from the DGP. Panel (a) presents the trend component as estimated by the Kalman filter, plotted against the actual trend. The corresponding figure for the smoothed estimate is given in panel (b). In both panels, the posterior median (black) as well 68th (solid) and 90th (dashed) percentile of the posterior density 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 all 26 variables. Both are normalized by subtracting the average volatility.

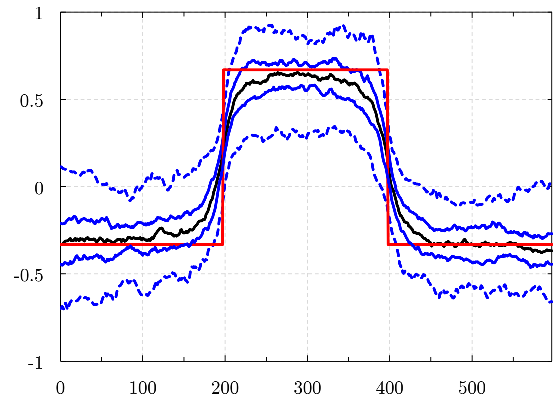
Figure B.2: Simulation Results II

Data-generating process (DGP) with two discrete breaks in the trend

(a) True vs. Estimated Trend (Filtered)



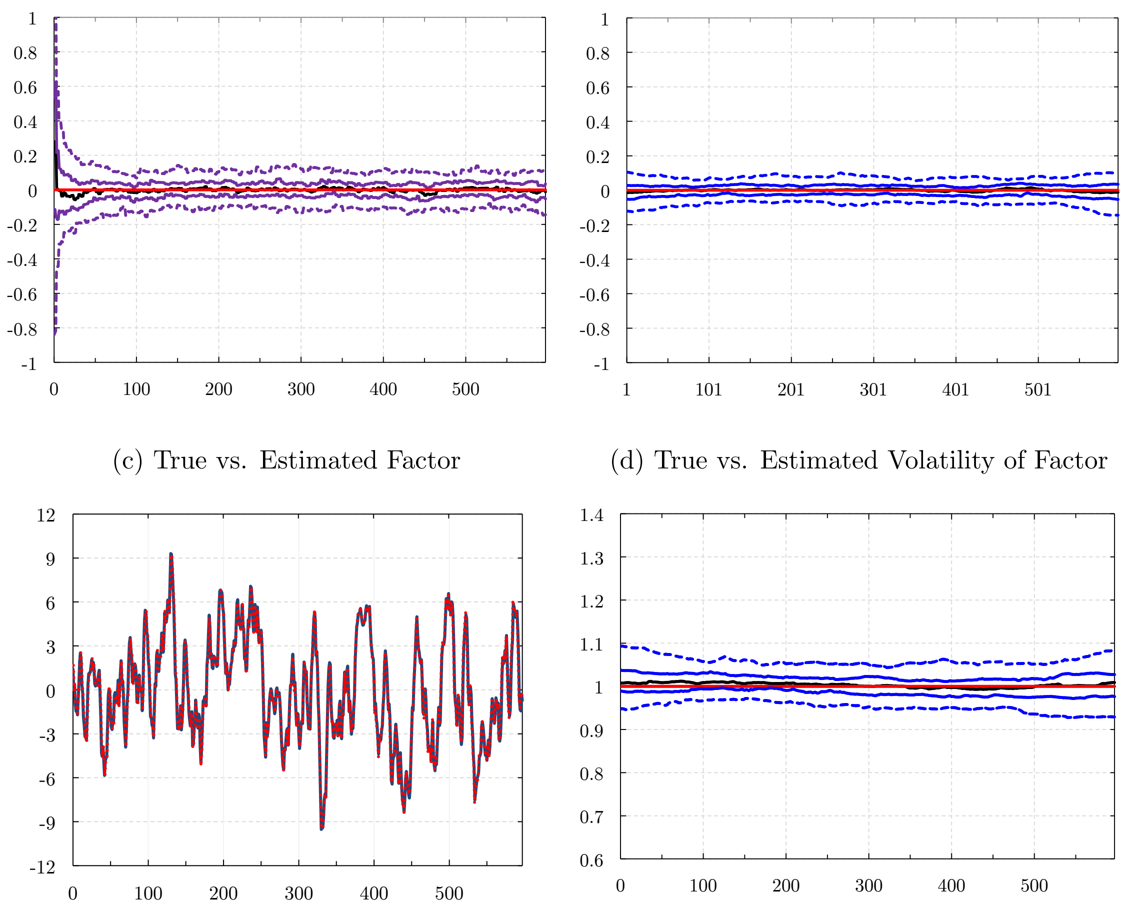
(b) True vs. Estimated Trend (Smoothed)



Note: The simulation setup is equivalent to the one in Figure B.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 posterior density. Panels (c) and (d) are omitted as they are very similar to Figure B.1.

Figure B.3: Simulation Results III

Data-generating process (DGP) without trend and without SV



(a) True vs. Estimated Trend (Filtered)

(b) True vs. Estimated Trend (Smoothed)

(c) True vs. Estimated Factor

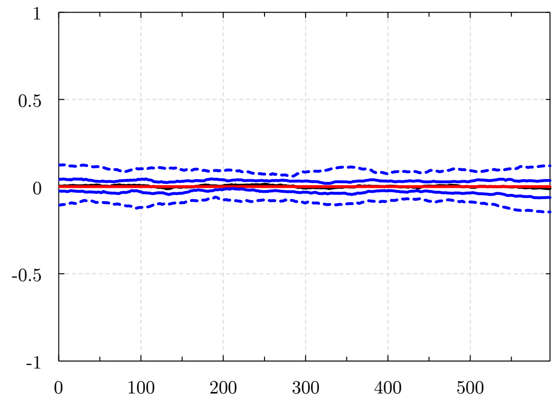
(d) True vs. Estimated Volatility of Factor

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 B.1. Again, we plot the filtered and smoothed median estimates of the trend with 68th and 90th percentiles of the posterior density 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.

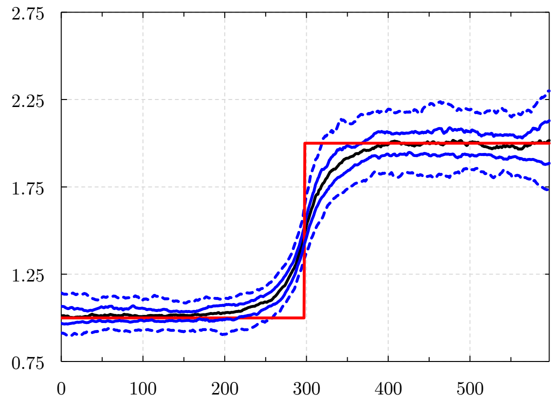
Figure B.4: Simulation Results IV

Data-generating process (DGP) without trend and discrete break in factor 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 B.1-B.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 posterior density 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 bands.

C Details on Estimation Procedure

C.1 Construction of the State Space System

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 vector $\tilde{y}_t$ be defined as

$$\tilde{y}_t = \begin{bmatrix} y^q_{1,t} \\ y_{2,t} - \rho_{2,1} y_{2,t-1} - \rho_{2,2} y_{2t-2} - \bar{\alpha_2} \\ \vdots \\ y_{n,t} - \rho_{n,1} y_{n,t-1} - \rho_{n,2} y_{n,t-2} - \bar{\alpha_n} \end{bmatrix},$$

where $\bar{\alpha}_i = \alpha_i(1 - \rho_{i,1} - \rho_{i,2})$, 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:

$$\begin{aligned} y_t &= HX_t + \tilde{\eta}_t, & \tilde{\eta}_t &\overset{iid}{\sim} N(0, \tilde{R}_t) \\ X_t &= FX_{t-1} + e_t, & e_t &\overset{iid}{\sim} N(0, Q_t) \end{aligned}$$

where the state vector is defined as $X_t = \left[\alpha_{1t}, \ldots, \alpha_{1t-4}, f_t, \ldots, f_{t-4}, u_{1t}, \ldots, u_{1t-4}\right]'$.

In the above state space system, after setting $\lambda_1 = 1$ for identification, the matrices of parameters H and F, are then constructed as follows:

$$H = \begin{bmatrix} H_q & H_q & H_q \\ \mathbf{0}_{(n-1)\times 5} & H_m & \mathbf{0}_{(n-1)\times 5} \end{bmatrix}$$

$$H_q = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 1 & \frac{2}{3} & \frac{1}{3} \end{bmatrix} \qquad H_m = \begin{bmatrix} \lambda_2 - \lambda_2\rho_{2,1} - \lambda_2\rho_{2,2} & 0 & 0 \\ \lambda_3 - \lambda_3\rho_{3,1} - \lambda_3\rho_{3,2} & 0 & 0 \\ & \vdots & \\ \lambda_n - \lambda_n\rho_{n,1} - \lambda_n\rho_{n,2} & 0 & 0 \end{bmatrix}$$

$$F = \begin{bmatrix} F_1 & \mathbf{0}_{5\times 5} & \mathbf{0}_{5\times 5} \\ \mathbf{0}_{5\times 5} & F_2 & \mathbf{0}_{5\times 5} \\ \mathbf{0}_{5\times 5} & \mathbf{0}_{5\times 5} & F_3 \end{bmatrix}$$

$$F_1 = \left[\begin{array}{c|c} 1 & \mathbf{0}_{1\times 4} \\ \hline I_4 & \mathbf{0}_{4\times 1} \end{array}\right] \qquad F_2 = \left[\begin{array}{cc|c} \phi_1 & \phi_2 & \mathbf{0}_{1\times 3} \\ \hline I_4 & & \mathbf{0}_{4\times 1} \end{array}\right] \qquad F_3 = \left[\begin{array}{cc|c} \rho_{1,1} & \rho_{1,2} & \mathbf{0}_{1\times 3} \\ \hline I_4 & & \mathbf{0}_{4\times 1} \end{array}\right]$$

Furthermore, the error terms are defined as

$$\begin{aligned}
\tilde{\eta}_t &= \begin{bmatrix} 0 & \eta_{2,t} & \dots & \eta_{n,t} \end{bmatrix}' \\
e_t &= \begin{bmatrix} v_{\alpha,t} & \mathbf{0}_{4\times 1} & \epsilon_t & \mathbf{0}_{4\times 1} & \eta_{1,t} & \mathbf{0}_{4\times 1} \end{bmatrix}'
\end{aligned}$$

with covariance matrices

$$\tilde{R}_t = \begin{bmatrix} 0 & \mathbf{0}_{1\times(n-1)} \\ \mathbf{0}_{(n-1)\times 1} & R_t \end{bmatrix},$$

where $R_t = diag(\sigma_{\eta_{2,t}}, \sigma_{\eta_{3,t}}, \dots, \sigma_{\eta_{n,t}})$,
and

$$Q_t = diag(\varsigma_{\alpha,1}, \mathbf{0}_{1\times 4}, \sigma_{\epsilon,t}, \mathbf{0}_{1\times 4}, \sigma_{\eta_{1,t}}, \mathbf{0}_{1\times 4}).$$

C.2 Details of the Gibbs Sampler

Let $\theta \equiv \{\lambda, \Phi, \rho, \varsigma_{\alpha_1}, \varsigma_\varepsilon, \varsigma_\eta\}$ be a vector that collects the underlying parameters. The model is estimated using a Markov Chain Monte Carlo (MCMC) Gibbs sampling algorithm in which conditional draws of the latent variables, $\{\alpha_{1t}, f_t\}_{t=1}^T$, the parameters, θ, 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.

C.2.0 Initialization

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

C.2.1 Draw latent variables conditional on model parameters and SVs

Obtain a draw $\{\alpha^j_{1t}, f^j_t\}_{t=1}^T$ from $p(\{\alpha_{1t}, f_t\}_{t=1}^T | \theta^{j-1}, \{\sigma^{j-1}_{\varepsilon,t}, \sigma^{j-1}_{\eta_{i,t}}\}_{t=1}^T, y)$.

This step of the algorithm uses the state space representation described above (Appendix C.1), and produces a draw from the entire state vector X_t 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 (2012), 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 3.2. Like Bai and Wang (2012), we initialise the Kalman Filter step from a normal distribution whose moments are independent of the model parameters, in particular $X_0 \sim N(0, 10^4)$.

C.2.2 Draw the variance of the time-varying GDP growth component

Obtain a draw $\varsigma^j_{\alpha_1}$ from $p(\varsigma_{\alpha_1} | \{\alpha^j_{1t}\}_{t=1}^T)$.

Taking the sample $\{\alpha^j_{1,t}\}_{t=1}^T$ drawn in the previous step as given, and posing an inverse-gamma prior $p(\varsigma_{\alpha_1}) \sim IG(S_{\alpha_1}, v_{\alpha_1})$ the conditional posterior of ς_{α_1} is also drawn inverse-gamma distribution. As discussed in Section 4.1, we choose the scale $S_{\alpha_1} = 10^{-3}$ and degrees of freedom $v_{\alpha_1} = 1$ for our baseline specification.

C.2.3 Draw the autoregressive parameters of the factor VAR

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

Taking the sequences of the common factor $\{f^{j-1}_t\}_{t=1}^T$ 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.

C.2.4 Draw the factor loadings

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

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 ρ^{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 Bai and Wang (2012).

C.2.5 Draw the serial correlation coefficients of the idiosyncratic components

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

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 can be obtained as $u_{i,t} = y_{i,t} - \lambda^{j-1} f_t^{j-1}$. 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.

C.2.6 Draw the stochastic volatilities

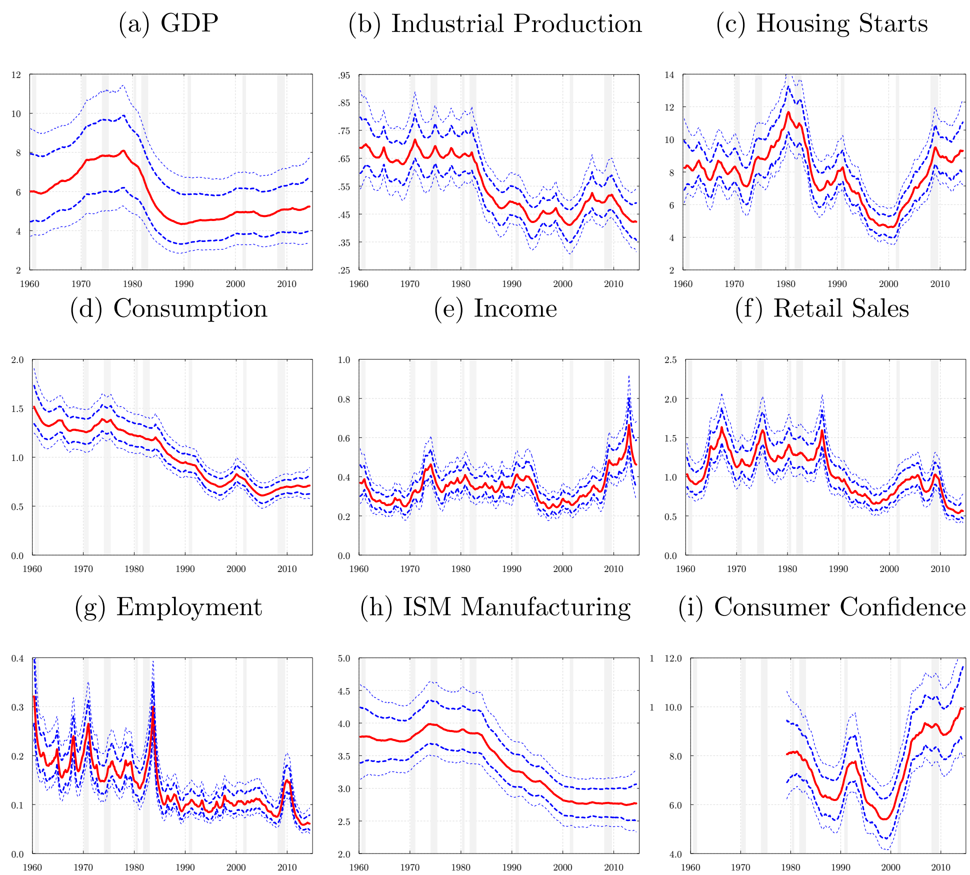
Obtain a draw of $\{\sigma_{\varepsilon,t}^j\}_{t=1}^T$ and $\{\sigma_{\eta_{i,t}}^j\}_{t=1}^T$ from $p(\{\sigma_{\varepsilon,t}\}_{t=1}^T|\Phi^{j-1}, \{f_t^{j-1}\}_{t=1}^T)$, and from $p(\{\sigma_{\eta_{i,t}}\}_{t=1}^T|\lambda^{j-1}, \rho^{j-1}, \{f_t^{j-1}\}_{t=1}^T, y)$ respectively.

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.

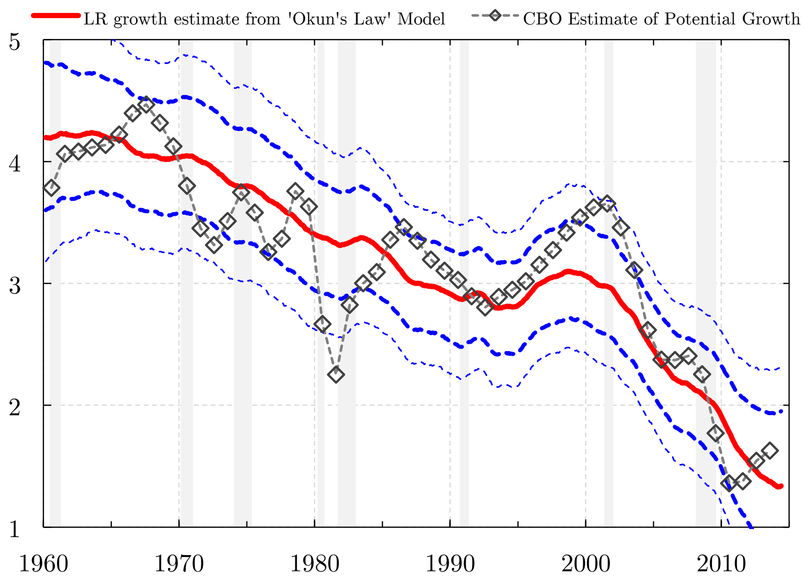
D Additional Figures

Figure D.1: Stochastic Volatility of Selected Idiosyncratic Components



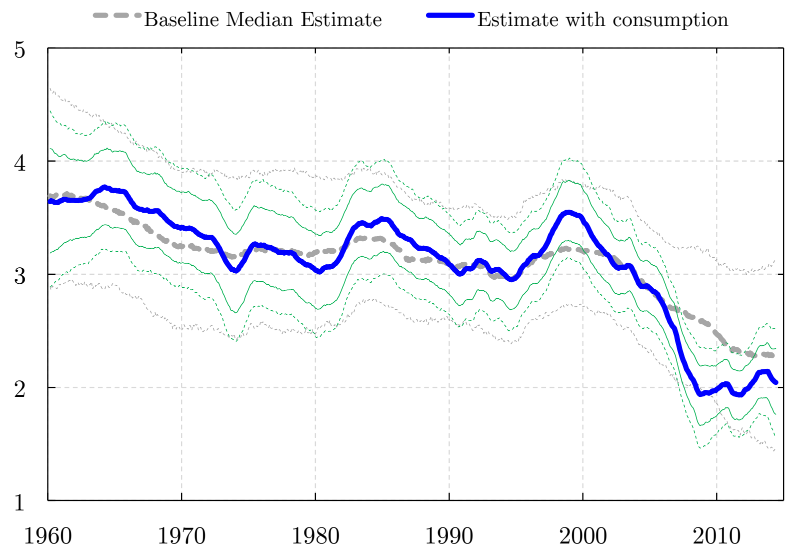
Note: Each panel presents the median (red), the 68% (dashed blue) and the 90% (dotted blue) posterior credible intervals of the idiosyncratic component of the idiosyncratic component of selected variables. Shaded areas represent NBER recessions. Similar charts for other variables are available upon request.

Figure D.2: Long-run GDP Growth Estimate using Unemployment Rate Only



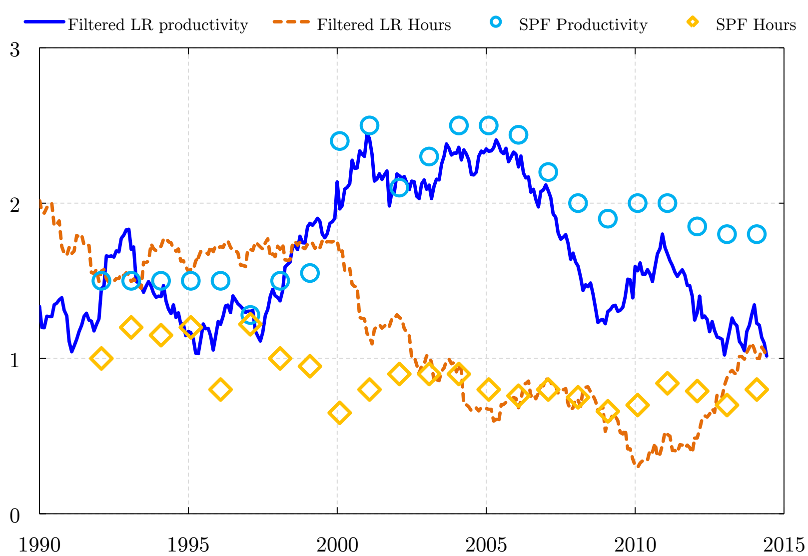
Note: The figure displays the posterior median estimate of long-run GDP growth for the model which uses unemployment as the single cyclical indicator, as discussed in Section 4.5. The 68% and 90% posterior credible intervals are shown. The combination of weak GDP growth and a strong decline in unemployment observed in the latter part of the sample drives down the estimate of long-run growth in this model, bringing it closer to the CBO's estimate.

Figure D.3: Long-run GDP Growth Estimates With And Without Including Consumption



Note: The figure displays the posterior median estimate of long-run GDP growth for the model with (solid blue) and without (dotted grey) including the series for consumption, as discussed in Section 5. The 68% and 90% posterior credible intervals are shown.

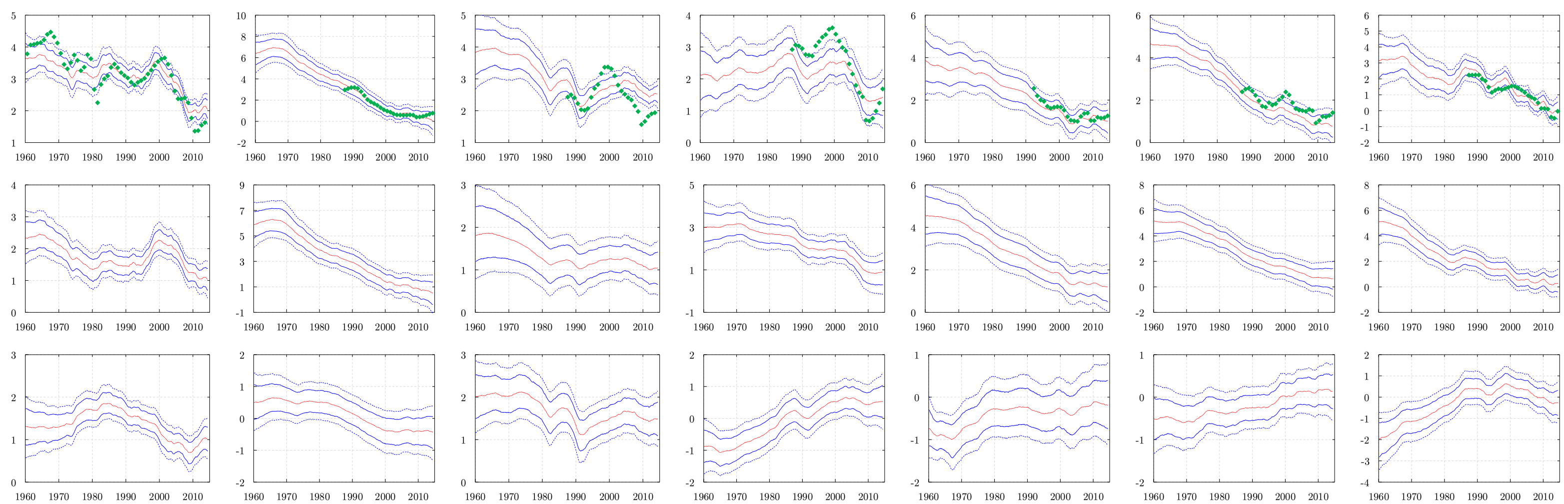
Figure D.4: Filtered Estimates of Long-Run Labor Productivity and Hours



Note: The figure displays the filtered estimates of the long-run labor productivity and hours component from the decompostion in Section 5, i.e. $\hat{z}_{t|t}$ and $\hat{h}_{t|t}$. For comparison, the corresponding forecasts from the SPF are plotted. The SPF forecast for total hours is obtained as the difference between the forecasts for real GDP and labour productivity.

Figure D.5: Posterior Estimates of Decomposition Exercise for G7 Economies

(a) United States (b) Japan (c) Canada (d) UK (e) Germany (f) France (g) Italy



Note: The upper panel for each country displays the overall long-run growth rate median estimate, together with the 68% and 90% posterior credible bands, as well as the CBO (United States) or OECD (remaining countries) estimate for comparison (green diamonds). The middle and lower panels show, respectively, the median estimates of z_t, the long-run growth rate of labor productivity, and h_t, the long-run growth rate of labor input, as well as the associated posterior credible bands.

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 occassions 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 datapoint was first published.

Table E.1:
DETAILED DESCRIPTION OF DATA SERIES

	Frequ.	Start Date	Vintage Start	Trans-formation	Publ. Lag	Data Code
Hard Indicators						
Real Gross Domestic Product	Q	Q1:1960	Dec 91	%QoQ Ann	26	GDPC1(F)
Real Industrial Production	M	Jan 60	Jan 27	% 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 Consumption Expenditures	M	Jan 60	Nov 79	% MoM	27	PCEC96(F)
Real Personal Income less Transfer Payments	M	Jan 60	Dec 79	% MoM	27	DSPIC96(F)
Real Retail Sales Food Services	M	Jan 60	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	Jan 60	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 60	May 55	% MoM	5	PAYEMS(F)
Civilian Employment (Household Survey)	M	Jan 60	Feb 61	% MoM	5	CE16OV(F)
Unemployed	M	Jan 60	Feb 61	% MoM	5	UNEMPLOY(F)
Initial Claims for Unempl. Insurance	M	Jan 60	Jan 00*	% MoM	4	LICM(H)

(Continues on next page)

Detailed Description of Data Series (Continued)

Soft Indicators						
Markit Manufacturing PMI	M	May 07	Jan 00*	-	-7	S111VPMM(H)[5] H111VPMM(H)
ISM Manufacturing PMI	M	Jan 60	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 Details About Forecast Evaluation

F.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 this paper with both time variation in the long-run growth of 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.[37]

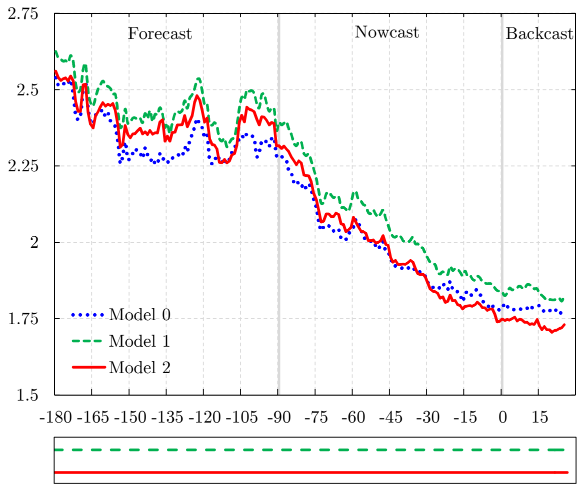
When comparing the three different models, we test the significance of any improvement of Models 1 and 2 relative to Model 0. This raises some important econometric complications given that (i) the three models are nested, (ii) the forecasts are produced using an expanding window, and (iii) the data 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, 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:Q1-2014:Q2) and the sample covering the recovery since the Great Recession (2009:Q2-2014:Q2).

[37] 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.

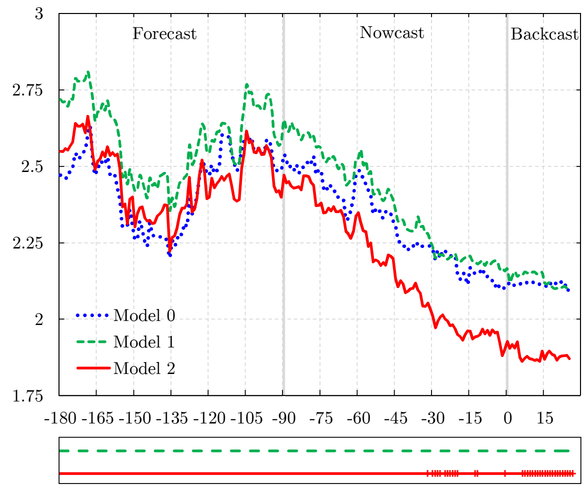
Figure F.1: Point Forecast Accuracy Evaluation

(a) Root Mean Squared Error

Full Sample: 2000:Q1-2014:Q2

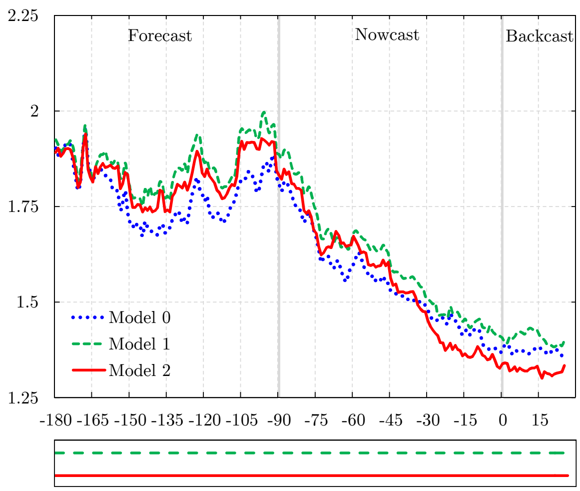


Recovery Sample: 2009:Q2-2014:Q2

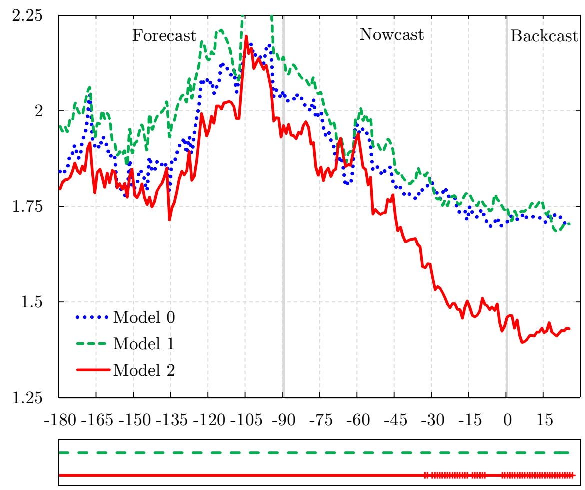


(b) Mean Absolute Error

Full Sample: 2000:Q1-2014:Q2



Recovery Sample: 2009:Q2-2014:Q2



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.

F.2 Point Forecast Evaluation

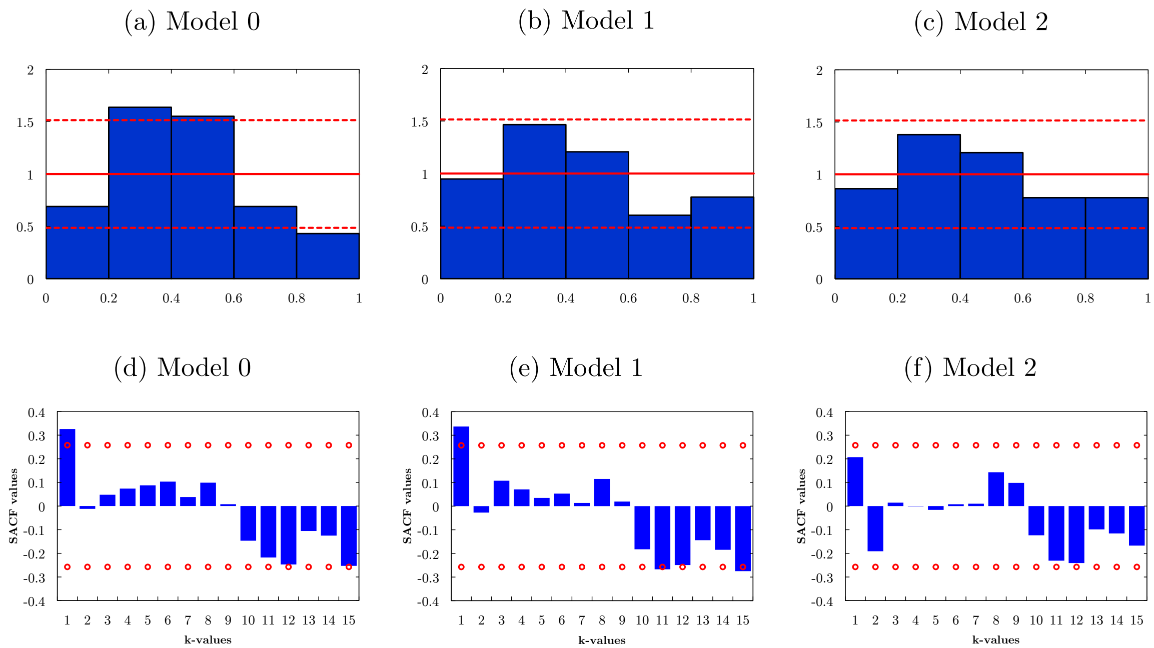
Figure F.1 shows the results of evaluating the posterior mean as point forecast. We use two criteria, the root mean squared error (RMSE) and the mean absolute error (MAE). As expected, both of these decline as the quarters advance and more information on monthly indicators becomes available (see e.g. Banbura et al., 2012). Both the RMSE and the MAE of Model 2 are lower than that of Model 0 starting 30 days before the end of the reference quarter, while Model 1 is somewhat worse overall. Although our gauge of significance indicates that these differences are not significant at the 10% level for the overall sample, 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 2014.

F.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 F.2 reports the probability integral transform (PITs) and the associated 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 F.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 hand 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 F.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,

Figure F.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.

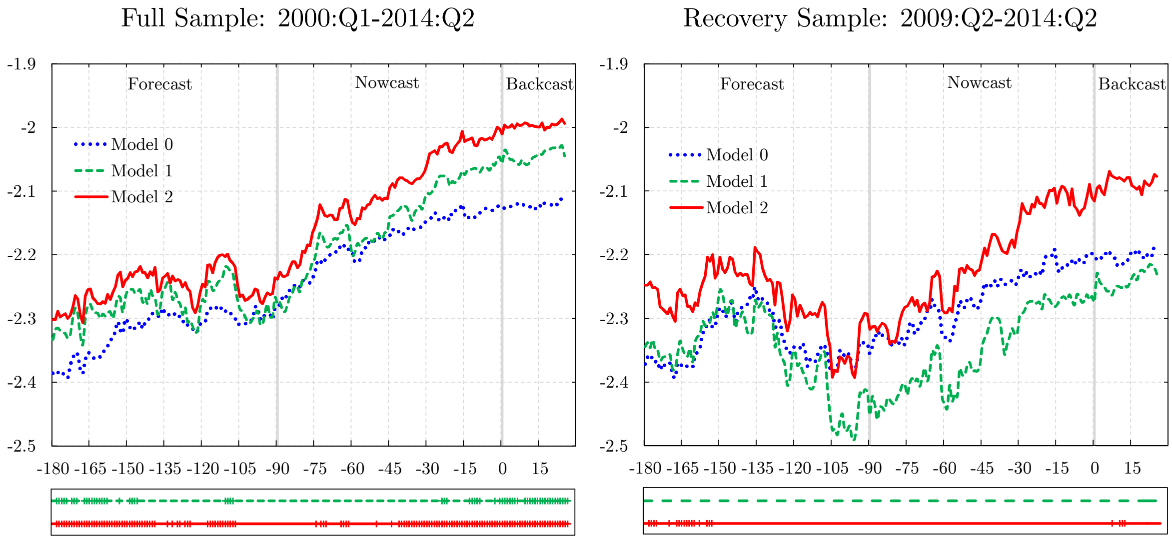
resulting in poorer density forecasts.

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 F.4).[38] 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

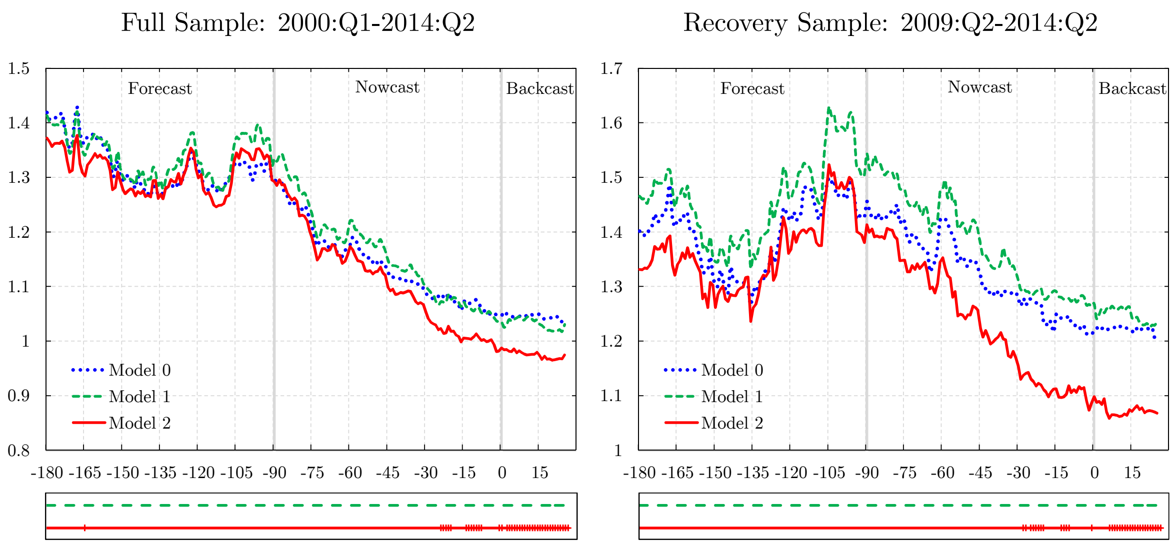
[38] As Gneiting and Ranjan (2011) show, integrating QS over the quantile spectrum gives the CRPS.

Figure F.3: Density Forecast Accuracy Evaluation

(a) Log Probability Score



(b) Continuous Rank Probability Score



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.

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).[39]

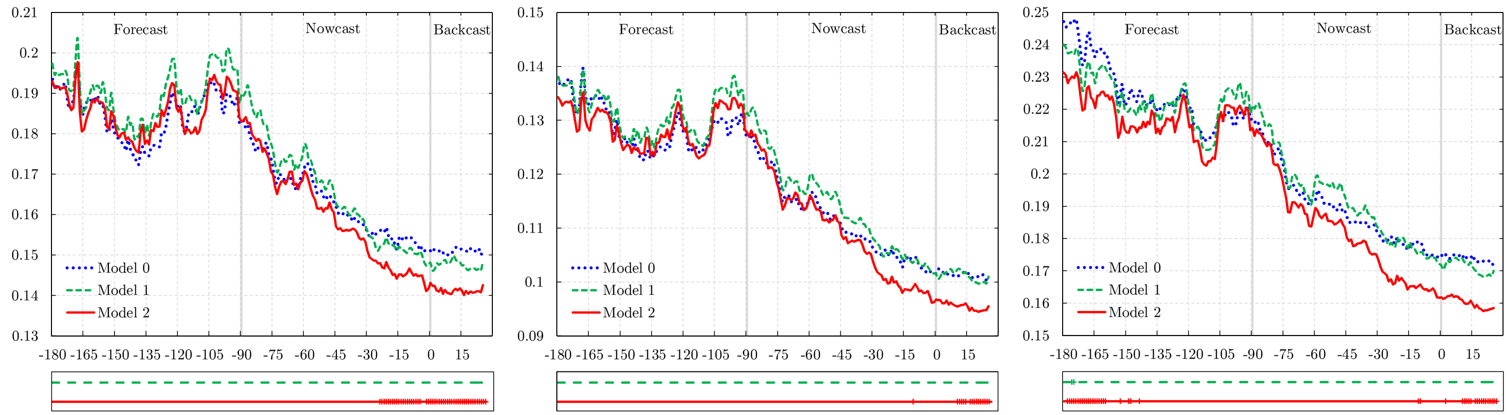
In sum, the results of the out-of-sample forecasting evaluation indicate that a model that allows for time-varying long-run GDP growth and SV produces short-run forecasts that are on average (over the full evaluation sample) either similar to or improve upon the benchmark model. The performance tends 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.

[39] The results are available upon request.

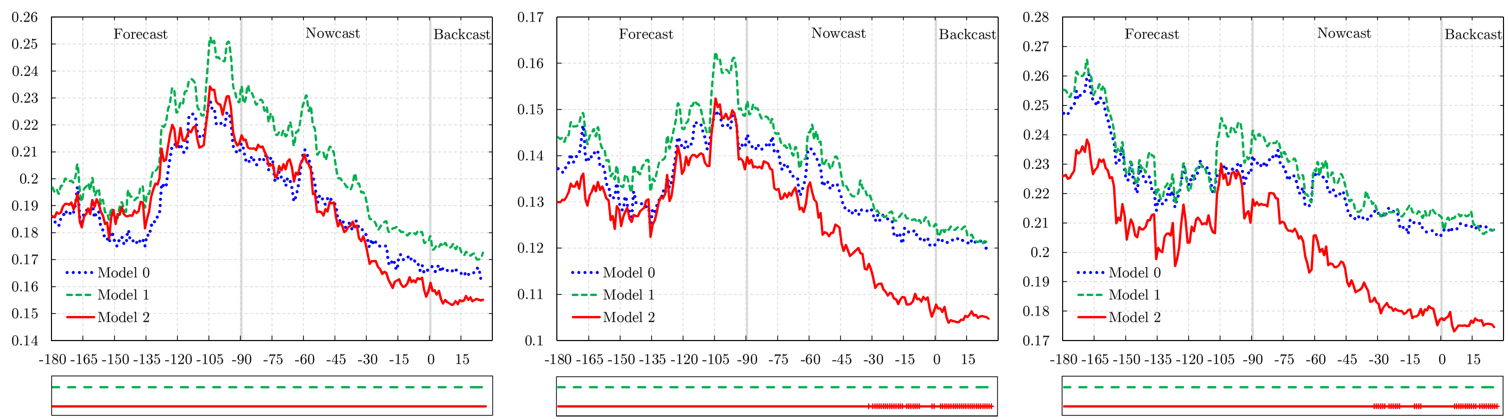
Figure F.4: Density Forecast Accuracy Evaluation: Quantile Score Statistics

Full Sample: 2000:Q1-2014:Q2

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



Recovery Sample: 2009:Q2-2014:Q2



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.

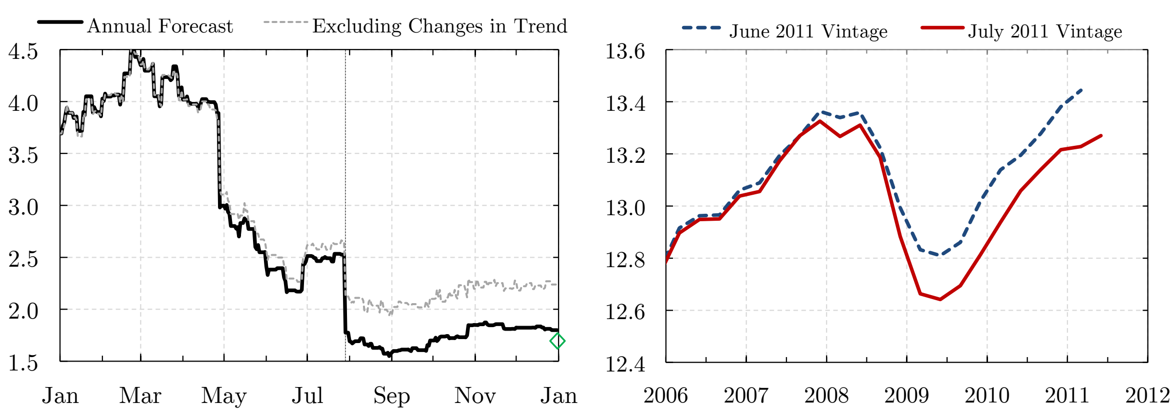
F.4 Case Study - The Decline of The Long-Run Growth Estimate in Mid-2011

Figure F.5 looks in more detail at the specific information that, in real time, led the model to reassess its estimate of long-run growth. While data since the middle of 2010 had already started shifting down the estimate, the publication by the Bureau of Economic Analysis of the annual revisions to the National Accounts in July 29th, 2011 were critical for the model's assessment. The vintages after that date show not just that the contraction in GDP during the recession was much larger than previously estimated, but even more importantly, it reduced the average growth rate of the subsequent recovery from 2.8% to about 2.4%. In short, by the end of July 2011 (well before structural break tests arrived at the same conclusion), enough evidence had accumulated about the weakness of the recovery to shift the model's estimate of long-run growth down by almost half a percentage point, very close to the final out-turn of 1.7%.

Figure F.5: Impact of July 2011 National Account Revisions

(a) Evolution of the Annual 2011 GDP Forecast

(b) National Accounts Revisions around July 2011



Note: Panel (a) shows the evolution of the annual GDP growth forecast for 2011 produced by the model. The light dashed line represents the counterfactual forecast that would result from shutting down fluctuations in long-run growth. The vertical line marks the release of the National Accounts revisions on 29th July 2011, and the green diamond marks the out-turn. Panel (b) shows the level of real GDP (in bn USD) before and after the revisions.