Business cycle dynamics after the Great Recession: An Extended Markov-Switching Dynamic Factor Model

Catherine Doz, Laurent Ferrara, Pierre-Alain Pionnier

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2020 – 03

Business cycle dynamics after the Great Recession: An Extended Markov-Switching Dynamic Factor Model

Catherine DOZ
Laurent FERRARA
Pierre-Alain PIONNIER

JEL Codes: C22, C51, E32, E37
Keywords: Markov-Switching Dynamic Factor Model (MS-DFM), Great Moderation, Great Recession, Turning-Point Detection, Macroeconomic Forecasting

Funded by a French government subsidy managed by the ANR under the framework of the Investissements d'avenir programme reference ANR-17-EURE-001

Business cycle dynamics after the Great Recession: An Extended Markov-Switching Dynamic Factor Model

Catherine DOZ[1]
Paris School of Economics (PSE) and University Paris 1 Panthéon-Sorbonne,
catherine.doz@univ-paris1.fr

Laurent FERRARA
SKEMA Business School – University Côte d'Azur
laurent.ferrara@skema.edu

Pierre-Alain PIONNIER
OECD, Statistics and Data Directorate
pierre-alain.pionnier@oecd.org

[1] We would like to thank Marcelle Chauvet, Guillaume Chevillon, Arnaud Dufays, Domenico Giannone, Sylvia Kaufmann, Siem-Jan Koopman, Josefine Quast, Luca Sala, Paul Schreyer, Dave Turner, and participants at the 2018 OxMetrics User Conference, the 2018 French Econometrics Conference, the 2019 Symposium of the Society for Nonlinear Dynamics and Econometrics (SNDE), the 2019 Conference of the International Association for Applied Econometrics (IAAE), the 2019 (3rd) Conference on Forecasting at Central Banks, the 2019 Conference on Real-Time Data Analysis, Methods and Applications, the 2019 CIRET/KOF/OECD/INSEE Workshop, and OECD, Namur University and Bocconi University seminars for helpful comments and suggestions.

Abstract: The Great Recession and the subsequent period of subdued GDP growth in most advanced economies have highlighted the need for macroeconomic forecasters to account for sudden and deep recessions, periods of higher macroeconomic volatility, and fluctuations in trend GDP growth. In this paper, we put forward an extension of the standard Markov-Switching Dynamic Factor Model (MS-DFM) by incorporating two new features: switches in volatility and time-variation in trend GDP growth. First, we show that volatility switches largely improve the detection of business cycle turning points in the low-volatility environment prevailing since the mid-1980s. It is an important result for the detection of future recessions since, according to our model, the US economy is now back to a low-volatility environment after an interruption during the Great Recession. Second, our model also captures a continuous decline in the US trend GDP growth that started a few years before the Great Recession and continued thereafter. These two extensions of the standard MS-DFM framework are supported by information criteria, marginal likelihood comparisons and improved real-time GDP forecasting performance.

Keywords: Markov-Switching Dynamic Factor Model (MS-DFM), Great Moderation, Great Recession, Turning-Point Detection, Macroeconomic Forecasting

JEL Codes: C22, C51, E32, E37

1. Introduction

Even though GDP is now widely recognised as an insufficient indicator to fully assess the health of an economy and how people are benefitting from economic growth (Stiglitz *et al.* 2018), it is also clear that recessions measured by GDP and other traditional macroeconomic indicators can have deep and long-lasting effects on people's (economic) well-being. For example, it is only at the end of 2017, nearly 10 years after the Great Recession of 2008-09, that the average unemployment rate in OECD countries fell back to its pre-crisis level. Moreover, the decline in subjective well-being has been especially visible in OECD countries most affected by the crisis (OECD 2013). Therefore, being able to detect recessions using the macroeconomic information that is available in real time, thus allowing to take adequate policy measures, is of key interest to policy makers.

The Great Recession, as well as the years following it, have been difficult times for macroeconomic forecasters. After making large forecast errors at the time of the Great Recession (Lewis and Pain 2015, An *et al.* 2018), the vast majority of macroeconomic forecasters have tended to systematically over-estimate GDP growth from 2011 to 2017, mainly by assuming long-term mean reversion in their forecasts, i.e. that economic growth would rapidly revert back to its long-term average (see Ferrara and Marsilli 2017 for an assessment of IMF forecasts). In addition, changes in macroeconomic volatility over the last decade further complicated the task of forecasters: after two decades of dampened macroeconomic fluctuations, referred to as the "Great Moderation" period (McConnell and Perez-Quiros, 2000), macroeconomic volatility increased during the Great Recession and apparently subsequently went down (Charles *et al.* 2018, Gadea-Rivas *et al.*, 2018).

In this paper, we consider small-scale Markov-Switching Dynamic Factor Models (MS-DFMs) as our starting point. These models, whose dynamics depends on whether the economy is in expansion or recession, are increasingly used for short-term forecasting and turning point detection (Chauvet and Potter 2013, Camacho *et al.* 2014, 2018). They have been introduced by Diebold and Rudebusch (1996) to simultaneously account for co-movement in macroeconomic time series and different dynamics during expansion and recession phases, two business cycle facts that were originally identified by Burns and Mitchell (1946). Chauvet and Piger (2008) and Hamilton (2011) later confirmed the very good empirical performance of small-scale MS-DFMs to track US recessions. They showed that MS-DFMs using the four monthly series advocated by the NBER Business Cycle Dating Committee (industrial production, real manufacturing trade and sales, real personal income excluding transfer payments and non-farm payroll employment) as observable variables are able to accurately replicate the NBER dating of US recessions in a fully transparent way. In addition, they showed that such models are able to identify business cycle phases faster than the NBER, especially around troughs. According to them, those small-scale MS-DFMs also compare favourably to the univariate Markov-Switching model of Hamilton (1989), which only extracts information from the US quarterly GDP series, and to the non-parametric multivariate algorithm of Harding and Pagan (2006), when it is used to extract information from the four monthly series advocated by the NBER for the dating of business cycles.

Against this background, our contribution to the literature is to put forward an Extended MS-DFM which is able to replicate the two recent empirical stylised facts observed since the Great Recession. The first one is the period of subdued GDP growth observed in most advanced economies since the Great Recession, or even slightly before. Deciding whether advanced economies have entered a period of secular stagnation and identifying causes for this stagnation is naturally of key interest for long-term forecasters and policy makers.

Nevertheless, Antolin-Diaz *et al.* (2017) also argue that, given the fact that standard DFM forecasts revert very quickly to the unconditional mean of GDP, accounting for a possible decline in long-term GDP growth may improve GDP forecasts even at very short horizons. Thus, we propose to integrate a time-varying GDP trend into a small-scale MS-DFM. As shown by Antolin-Diaz *et al.* (2017), the explicit modelling of time variation in long-run GDP growth is likely to deliver substantial gains in out-of-sample forecasting performance for the US after the Great Recession. To account for the second stylized fact (i.e. changing macroeconomic volatility), we allow for switches in volatility in the state equation of the model, according to a Markov chain that differs from the one we use for the conditional mean. Our approach can be linked to that of Chauvet and Su (2014) who put forward a univariate model with changes in mean and volatility governed by Markov-Switching processes, as well as structural breaks.

Based on US data, we show that the introduction of Markov-Switching volatility in the standard MS-DFM framework is supported by statistical criteria (DICs and marginal likelihoods) and improves the detection of turning points after the mid-1980s. The Extended MS-DFM also accurately detects the Great Recession in advance to NBER announcements. The timely detection that the economy has switched to a recession regime allows an improvement of real-time GDP forecasts during the Great Recession as compared to those stemming from a linear DFM incorporating the same amount of information. The introduction of a time-varying long-term GDP growth rate is most helpful to improve real-time GDP forecasts after the Great Recession.

As a by-product of the estimation of our Extended MS-DFM on US data, we also get two interesting results. First, we note a decline in the US long-run GDP growth rate of about one p.p. per year as compared to the early 2000s, half of it occurring before the Great Recession. Such decline in long-run US GDP growth has been empirically documented by Antolin-Diaz *et al.* (2017) and rationalised by Fernald *et al.* (2017) who highlight the role of total factor productivity and labour force participation. Second, we conclude that the Great Recession did not put an end to the Great Moderation. According to our model, the Great Recession, beyond its effect on GDP growth, was characterised by a temporary increase in macroeconomic volatility in 2008-09, but we observe a return to a low-volatility environment from 2010 onwards. This result is important for the calibration of macroeconomic models and for the forecasting of future US recessions. It is also in line with recent economic research (Charles *et al.*, 2018, Gadea-Rivas *et al.*, 2018).

The present paper is organised as follows. In Section 2, we review the related literature. In Section 3, we extend the standard MS-DFM framework to account for both switches in macroeconomic volatility and time variation in long-term GDP growth. In Section 4, we describe a Bayesian (Gibbs-sampling) methodology to estimate this model. In Section 5, we provide in-sample estimation results based on US data from 1970 to 2017. In Section 6, we evaluate the real-time out-of-sample properties of our model.

2. Related literature

Initially introduced in the literature by Diebold and Rudebusch (1996), Markov-Switching Dynamic Factor Models (MS-DFMs) have two attractive features: they allow to simultaneously account for co-movement in macroeconomic time series, and for different dynamics during expansion and recession phases. Chauvet (1998) and Kim and Nelson (1998) later proposed maximum likelihood and Bayesian estimation strategies for these models.

In this paper, we put forward an Extended MS-DFM that also incorporates a time-varying long-run GDP growth and Markov-Switching volatility. Our model can be seen as a non-linear extension of the model proposed by Antolin-Diaz *et al.* (2017), who include a time-varying long-run GDP growth rate in an otherwise linear DFM model. Our main difference with them is the inclusion of Markov-Switching features in the state equation of our model to disentangle expansion from recession phases, as well as high-volatility from low-volatility regimes.

Like Antolin-Diaz *et al.* (2017) and Marcellino *et al.* (2016), we account for time-variation in macroeconomic volatility, but whereas they rely on a stochastic volatility specification, we rely on a Markov-Switching volatility specification. The evidence provided by Sensier and van Dijk (2004) supports this specification. Indeed, they document that the vast majority (80%) of US macroeconomic variables experienced a reduction in (conditional) volatility after 1980 and that these volatility changes are better characterised by instantaneous breaks than by gradual changes. Stock and Watson (2002) also conclude that the Great Moderation episode is better accounted for by a sharp break in volatility rather than a gradual decline, thus reflecting a reduction in the size of shocks hitting the economy more than a change in the propagation mechanism of these shocks.

Our model is closely connected to the one proposed by Eo and Kim (2016) who relax the assumption that all recessions and expansions are alike in a univariate Markov-Switching model of the business cycle. But whereas Eo and Kim (2016) allow for a fully flexible evolution of the regime-specific mean growth rates over time, with random walks on the long-run GDP growth rate and the regime-specific mean growth rates, we relate evolutions in regime-specific mean growth rates to evolutions in macroeconomic volatility, as proposed by McConnell and Perez-Quiros (2000), Giordani *et al.* (2007) and Bai and Wang (2011). It is also connected to the model proposed by Chauvet and Su (2014), who include independent Markov-Switching processes for the mean and the variance of GDP growth, and add a structural break in GDP growth, whereas we put forward a time-varying long-run GDP growth rate.

In this paper, we consider a larger information set than Eo and Kim (2016) or Chauvet and Su (2014), and include the four monthly variables advocated by the NBER in addition to quarterly GDP, the only variable that they consider. The fact that we consider a multivariate framework also distinguishes the present model from the univariate models of McConnell and Perez-Quiros (2000), Giordani *et al.* (2007) and Bai and Wang (2011). In addition, extracting a common factor from multiple series also makes it less important to explicitly account for outliers in individual observed series, as advocated by Giordani *et al.* (2007). Note that our information set could easily be expanded to include additional monthly or quarterly indicators, which makes our model suitable for nowcasting GDP growth.

Similarly to us, Eo and Morley (2017) also consider two types of recessions and allow for a break in trend GDP growth in 2006Q1. Nevertheless, they do not relate the two recession types to macroeconomic volatility but distinguish L-shaped recessions, with a permanent effect on the level of GDP, and U-shaped recessions, with only a transitory

effect on the level of GDP. Like Eo and Kim (2016), their information set only includes US quarterly GDP. Based on this specification, they conclude that the Great Recession was U-shaped and that the US economy fully recovered from it in 2014.

Antolin-Diaz *et al.* (2017), Eo and Kim (2016) and Eo and Morley (2017) identify a decrease in US trend GDP growth which started before the Great Recession, but its precise timing and magnitude differ. Antolin-Diaz *et al.* (2017) conclude that the US long-run GDP growth has declined from a peak of 3.5% per year around 2000 to about 2.0% per year in 2015. Eo and Kim (2016) identify a continuous decrease in US trend GDP growth from 3.6% per year in the 1950s to 2.0% per year around 2010, with no noticeable peak around 2000. [2] Finally, Eo and Morley (2017) report a break of -0.52% per quarter (or - 2.06% per year) in US trend GDP growth in 2006Q1. The differences between these three estimates may be related to differences in information sets (16 monthly and quarterly series for Antolin-Diaz *et al.* (2017), US quarterly GDP growth for both Eo and Kim (2016) and Eo and Morley (2017)), differences in the modelling of trend GDP growth (continuous variation for both Antolin-Diaz *et al.* (2016) and Eo and Kim (2016), discrete break for Eo and Morley (2017)), and differences in the modelling of cyclical fluctuations.

Table 1 summarises how the present paper relates to the existing literature.

Table 1. Comparison with the existing literature

	Multivariate framework (factor model)	Markov-Switching Intercept	Markov-Switching volatility of shocks	Stochastic volatility	Time-varying long-term GDP growth rate
Antolin-Diaz *et al.* (2017)	X			X	X
Bai and Wang (2011)		X	X		
Chauvet and Su (2014)		X	X		
Diebold and Rudebusch (1996)	X	X			
Eo and Kim (2016)		X		X	X
Giordani *et al.* (2007)[3]		X	X		X
Kim and Nelson (1998)	X	X			
Marcellino *et al.* (2016)	X			X	
McConnell and Perez-Quiros (2000)		X	X		
This paper	**X**	**X**	**X**		**X**

[2] Eo and Kim (2016) actually report quarterly figures.

[3] Giordani *et al.* (2007) actually consider the dynamics of industrial production (IP) in G7 countries. Thus, they introduce time variation in the long-term growth rate of IP, not GDP, but do it in a very similar way as in the present paper and Antolin-Diaz *et al.* (2017).

3. Specification of the Extended MS-DFM

The factor model with Markov-Switching mean was introduced by Diebold and Rudebusch (1996) and estimated in a Bayesian way (Gibbs Sampling) by Kim and Nelson (1998). In the model we put forward in this paper, it is assumed that both the intercept and the volatility of shocks in the state equation (factor dynamics) are governed by two independent Markov-Switching processes. This specification can be thought as a generalisation of McConnell and Perez-Quiros (2000) to a multivariate framework. Indeed, McConnell and Perez-Quiros (2000) only considered US quarterly GDP as observable variable, whereas we consider the four monthly series explicitely mentioned by the business cycle Dating Committee of the NBER, in addition to quarterly GDP.

Let us first consider a general state-space model where, for the time being, all the *n* observable variables y_{it} are available at monthly frequency. This model is given by the two following equations:

Measurement equations (for i = 1,...,n): $\Delta y_{it} = a_{i,t} + \gamma_i(L)\Delta c_t + u_{it}$

State equations: $$\begin{cases} a_{i,t} = a_{i,t-1} + \sigma_{a,i} \cdot \eta_t^{a,i} \quad ; \quad \eta_t^{a,i} \sim IID\ N(0,1) \\ \phi(L)\Delta c_t = \mu_{S_t,V_t} + \underbrace{\sqrt{1 + h \cdot V_t} \cdot \sigma_c \cdot \eta_t^c}_{\substack{Variance = \sigma_c^2 \text{ in the low-volatility regime,} \\ \text{and } (1+h)\cdot \sigma_c^2 \text{ in the high-volatility regime}}} \quad ; \quad \eta_t^c \sim IID\ N(0,1) \\ \psi_i(L)u_{it} = \sigma_i \cdot \varepsilon_{it} \quad ; \quad \varepsilon_{it} \sim IID\ N(0,1) \end{cases}$$

In the measurement equations, the Δc_t component captures common fluctuations across observable variables and the $a_{i,t}$ variables capture potential low-frequency variations in mean growth rates, such as the slow decline in the GDP growth rate, as in Giordani *et al.* (2007) and Antolin-Diaz *et al.* (2017). This random walk that we introduce in the data generating process of the GDP growth rate should only be considered as a convenient modelling simplification. It allows to capture the low-frequency component of the time series and remains innocuous over a limited period of time. Our approach can also be related to the one of Stock and Watson (2012) who subtract a local mean to the US GDP growth rate before including it in their model. In our case, this adjustment is made within the model.

In the state equations, the $a_{i,t}$ variables are modelled as independent random walks, the common component Δc_t follows an autoregressive process with Markov-Switching intercept and Markov-Switching variance, and the error terms u_{it} are modelled as independent autoregressive processes. Thus, we allow the error terms of the measurement equations to be autocorrelated, but not cross-correlated. The fact that the business cycle, proxied by the common factor, directly depends on the volatility is based on the literature pointing out that second moments are useful to predict macroeconomic fluctuations (see McConnell and Perez-Quiros 2000, Bai and Wang 2011, and Chauvet *et al.* 2015).

In addition, S_t and V_t are two independent first-order Markov-Switching processes, each with only two possible states (0 and 1). Transition probabilities are given by:

$$p(S_t = 0|S_{t-1} = 0) \equiv PrS_{00} \quad ; \quad p(S_t = 1|S_{t-1} = 1) \equiv PrS_{11}$$
$$p(V_t = 0|V_{t-1} = 0) \equiv PrV_{00} \quad ; \quad p(V_t = 1|V_{t-1} = 1) \equiv PrV_{11}$$

And the intercept of the state equation is such that:

$$\mu_{S_t,V_t} \equiv \mu_{00} + \underbrace{\mu_{01}}_{<0} \cdot V_t + \underbrace{\mu_{10}}_{>0} \cdot S_t + \underbrace{\mu_{11}}_{>0} \cdot S_t \cdot V_t$$

where

μ_{00}: Intercept in the low-volatility recession regime

$\mu_{00} + \underbrace{\mu_{01}}_{<0}$: Intercept in the high-volatility recession regime

$\mu_{00} + \underbrace{\mu_{10}}_{>0}$: Intercept in the low-volatility expansion regime

$\mu_{00} + \underbrace{\mu_{01}}_{<0} + \underbrace{\mu_{10}}_{>0} + \underbrace{\mu_{11}}_{>0}$: Intercept in the high-volatility expansion regime

In the empirical part of this paper, we will use a mix of quarterly and monthly data in the information set: the quarterly GDP growth rate and monthly growth rates of industrial production, real manufacturing trade and sales, real personal income excluding transfer payments, and non-agricultural civilian employment. Let us first clarify how the model can accommodate this mix of monthly and quarterly observable variables. Quarterly GDP Δy_{1t}^q is observed once every three months whereas monthly variables Δy_{jt}^m are observed every month. This is something that the Kalman filter can easily accommodate. The key to relate quarterly GDP to the underlying monthly state variables is to approximate the quarterly GDP growth rate as a weighted average of current and past monthly GDP growth rates, a solution that has been popularised by Mariano and Murasawa (2003).

The five measurement equations can thus be rewritten as follows:

$$\begin{cases} \Delta y_{1t}^q = \left(\frac{1}{3}a_{1,t}^q + \frac{2}{3}a_{1,t-1}^q + a_{1,t-2}^q + \frac{2}{3}a_{1,t-3}^q + \frac{1}{3}a_{1,t-4}^q\right) + \gamma_1^q(L)\left(\frac{1}{3}\Delta c_t + \frac{2}{3}\Delta c_{t-1} + \Delta c_{t-2} + \frac{2}{3}\Delta c_{t-3} + \frac{1}{3}\Delta c_{t-4}\right) \\ \qquad + \left(\frac{1}{3}u_{1,t}^q + \frac{2}{3}u_{1,t-1}^q + u_{1,t-2}^q + \frac{2}{3}u_{1,t-3}^q + \frac{1}{3}u_{1,t-4}^q\right) \\ \Delta y_{jt}^m = a_{j,t}^m + \gamma_j^m(L)\Delta c_t + u_{j,t}^m \end{cases}$$

Then, we assume that quarterly GDP and the first three monthly variables (industrial production, real manufacturing trade and sales, and real personal income excluding transfer payments) are only linked contemporaneously to the business cycle (i.e. to the underlying factor). In other words, we assume that:

$$\gamma_1^q(L) = \gamma_{10}^q$$

$$\gamma_j^m(L) = \gamma_{j0}^m \quad \text{for } j = 1 \dots 3$$

Besides, as in any factor model, we have to impose an identifying restriction (otherwise, the factor would be defined up to a multiplicative constant). Here, we impose the following restriction:

$$\gamma_{10}^q \equiv 1$$

For the fourth monthly variable (non-agricultural civilian employment), we include a richer lag structure in order to take into account that employment may be lagging, as in Stock and Watson (1989):

$$\gamma_4^m(L) = \gamma_{40}^m + \gamma_{41}^m L + \gamma_{42}^m L^2 + \gamma_{43}^m L^3$$

Next, we follow Chauvet and Hamilton (2005) in specifying an AR(1) dynamics for the idiosyncratic shocks and the underlying factor, that is:

$$\psi_1^q(L) = 1 - \psi_{11}^q L$$

$$\psi_j^m(L) = 1 - \psi_{j1}^m L \quad \text{for } j = 1 \dots 4$$

$$\phi(L) = 1 - \phi_1 L$$

Lastly, we assume that, in our information set, only GDP growth exhibits low-frequency fluctuations, captured by $a_{1,t}^q$, which is modelled as a random walk without drift. Since all observable variables are demeaned prior to model estimation, there is no need to add constant terms in the measurement equations. All monthly variables, expressed as monthly growth rates, are assumed to be driven by cyclical fluctuations captured by the underlying factor Δc_t and the idiosyncratic shocks $u_{j,t}^m$. This implies $a_{j,t}^m = 0$ for $j = 1 \dots 4$. We thus limit the parameterisation of our model and only explicitly account for low-frequency fluctuations of our main variable of interest, which is GDP growth. Antolin-Diaz *et al.* (2017) follow a similar modelling strategy and run simulation experiments showing that not including a time-varying trend for the other variables is innocuous for the estimation of the GDP trend as long as persistence is allowed for in the idiosyncratic components of all variables. This is precisely the role of the $\psi_j^m(L)$ polynomials in our model, for which we do not rule out the possibility of having unit roots.

Additional details on the state-space representation of the model are available in Annex E.

4. Data and estimation strategy

4.1. Data sources

As previously mentioned, our information set contains five variables: quarterly real GDP and four monthly variables (industrial production, real manufacturing trade and sales, real personal income excluding transfer payments and non-agricultural civilian employment). These are the variables advocated by the NBER Business Cycle Dating Committee (BCDC). Note that following Chauvet and Hamilton (2005), we rely on civilian employment rather than payroll employment[4] and show the advantage of doing so in Annex B. The US quarterly, real and seasonally adjusted GDP is extracted from the ALFRED database maintained by the Federal Reserve Bank of St Louis. The corresponding quarterly vintages are available from December 1991 onwards. The four monthly series in the information set are extracted from the FRED-MD database maintained by the Federal Reserve Bank of St Louis (McCracken and Ng 2016).[5] These monthly series are all seasonally adjusted. The corresponding monthly vintages are available from August 1999 onwards. In the FRED-MD database available for month M, industrial production and non-agricultural civilian employment are typically available up until month (M-1), real personal income excluding transfer payments up until month (M-1) or (M-2), and real manufacturing trade and sales up until month (M-2) or (M-3).[6] In the following, only real-time samples, corresponding to monthly releases of the FRED-MD database, are used.

4.2. Estimation strategy

We estimate the model in a Bayesian way and rely on Gibbs Sampling. We see two main advantages of the Bayesian estimation strategy. First, it is a modular estimation technique, allowing to add or remove building blocks in the model and compare model specifications with each other. Second, it simplifies the inference on the Markov-Switching variables S_t and V_t because it allows conditioning on the underlying factor and treat it as if it was an observed monthly variable. As a consequence, the inclusion of both quarterly and monthly variables in the information set does not complicate the inference on S_t and V_t. This is a key advantage as compared to the strategy put forward by Camacho *et al.* (2018) who directly rely on the variables in the information set to estimate the underlying Markov-Switching variables. When the variables in the information set have different frequencies, their distribution potentially depends on many lags of the Markov-Switching variables, which generates a curse of dimensionality problem which Camacho *et al.* (2018) solve at the price of approximations. None of these approximations are needed with our Bayesian estimation strategy.

---

[4] Payroll employment is based on a survey of business establishments whereas civilian employment is based on a household survey. As noted by Chauvet and Hamilton (2005), payroll employment only includes job creation and destruction with a lag and does not include self-employment or off-the-book employment, which can delay the detection of a recovery. Moreover, it double counts jobs if a person changes job within a payroll survey reference period, which can overestimate employment around peaks. Lastly, payroll employment undergoes substantial revisions over time whereas civilian employment does not get revised.

[5] GDP, industrial production, trade and sales, personal income and employment are respectively available using the following codes: GDPC1, INDPRO, CMRMTSPLx, W875RX1, CE16OV.

[6] Real personal income excluding transfer payments and real manufacturing trade and sales are not available with a fixed lag in the FRED-MD database. The corresponding lags depend on the year and the month which are considered.

Our Gibbs Sampling algorithm includes the following four main blocks, which are sequentially iterated until convergence. Subsequently, we use the following notation for all time series: $\widetilde{\alpha_{1..T}} = \{\alpha_1, \dots, \alpha_T\}$.

1. Draw the state vector $\widetilde{\alpha_{1..T}}$ conditional on $\widetilde{\Delta y^*_{1..T}}$, $\widetilde{S_{1..T}}$, $\widetilde{V_{1..T}}$ and the model parameters, based on the sequential Kalman filter with diffuse initialisation of Koopman and Durbin (2000, 2003) and the simulation smoother of Durbin and Koopman (2002). The sequential Kalman filter considers the series in the information set one by one when updating estimates of the state vector. Beyond allowing some savings in computing time, it also simplifies diffuse initialisation as compared to the multivariate Kalman filter. Moreover, both the Kalman filter and the simulation smoother efficiently deal with missing observations, which is crucial in the present context where the variables in the information set are released at different dates (ragged-edge sample) and have different frequencies.
2. Draw $\widetilde{S_{1..T}}$ conditional on $\widetilde{\alpha_{1..T}}$, $\widetilde{V_{1..T}}$ and model parameters, based on Hamilton's (1989) filter.
3. Draw $\widetilde{V_{1..T}}$ conditional on $\widetilde{\alpha_{1..T}}$, $\widetilde{S_{1..T}}$ and model parameters, based on Hamilton's (1989) filter.
4. Sequentially draw the different model parameters, based on $\widetilde{\Delta y^*_{1..T}}$, $\widetilde{\alpha_{1..T}}$, $\widetilde{S_{1..T}}$, $\widetilde{V_{1..T}}$ and the other model parameters.

Additional details on the estimation strategy are available in Annexes C and D.

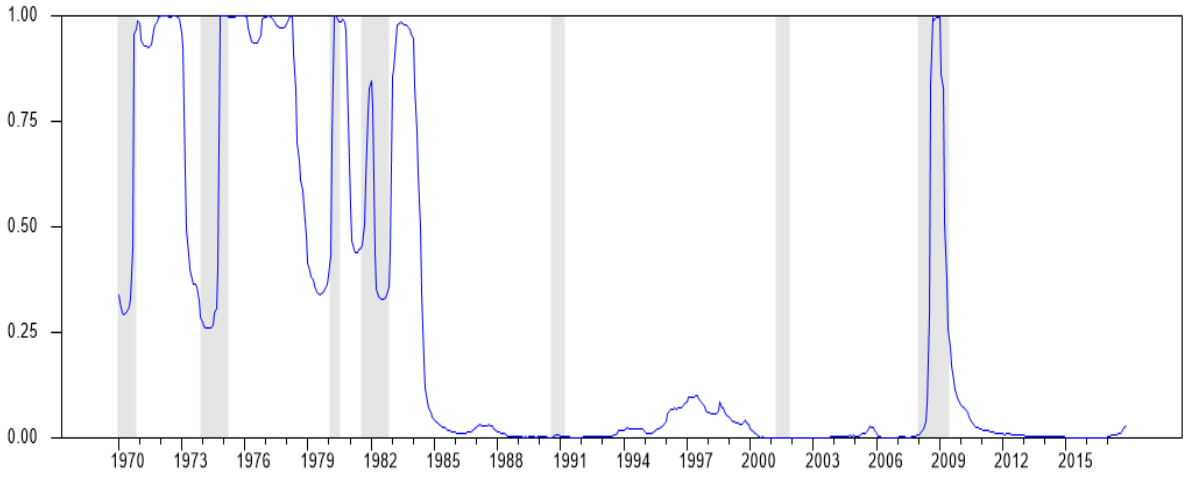
5. In-sample results from the Extended MS-DFM specification

In this section, we present in-sample results estimated by using data from January 1970 to December 2017.

5.1. Different macroeconomic volatility regimes since 1970

The inclusion of switches in volatility in the state equation of the model enables to distinguish four periods in the sample, as shown in Figure 1: a period of high macroeconomic volatility from 1970 to 1984, a period of low volatility from 1984 to 2007 corresponding to the Great Moderation, a second period of high macroeconomic volatility corresponding to the Great Recession (2007-2009), and a period of low volatility from the end of the Great Recession to the end of the sample (2009-2017). Looking at probabilities of being in a high-volatility regime (Figure 1), the transitions between the four periods are quite sharp. From Figure 1, we also conclude that the increase in volatility at the time of the Great Recession was only temporary and did not put an end to the Great Moderation, which is in line with Gadea-Rivas *et al.* (2018) and Charles *et al.* (2018). We also identify more minor fluctuations of the probability of being in a high-volatility state during the first period (1970-1984), something which is also noticed by Antolin-Diaz *et al.* (2017) who rely on a stochastic volatility specification.

Figure 1. Extended MS-DFM: Smoothed probability of being in a high-volatility regime



Estimation sample: 1970M01-2017M12. Data vintage: 2017M12. Estimation based on 1 5000 draws of the Gibbs Sampler (the first 5 000 are discarded).

5.2. Improved detection of recessions when allowing for different characteristics of expansions and recessions between volatility regimes

As shown in Figure 2, allowing for switches in volatility in the state equation and for different expansion and recession characteristics across volatility regimes also helps to better capture the two recessions that occurred during the Great Moderation period. This finding is in line with McConnell and Perez-Quiros (2000) who showed using a univariate MS-model that these two modelling features improved the detection of turning

points after the mid-1980s. We extend their findings to a multivariate context and a more recent sample.

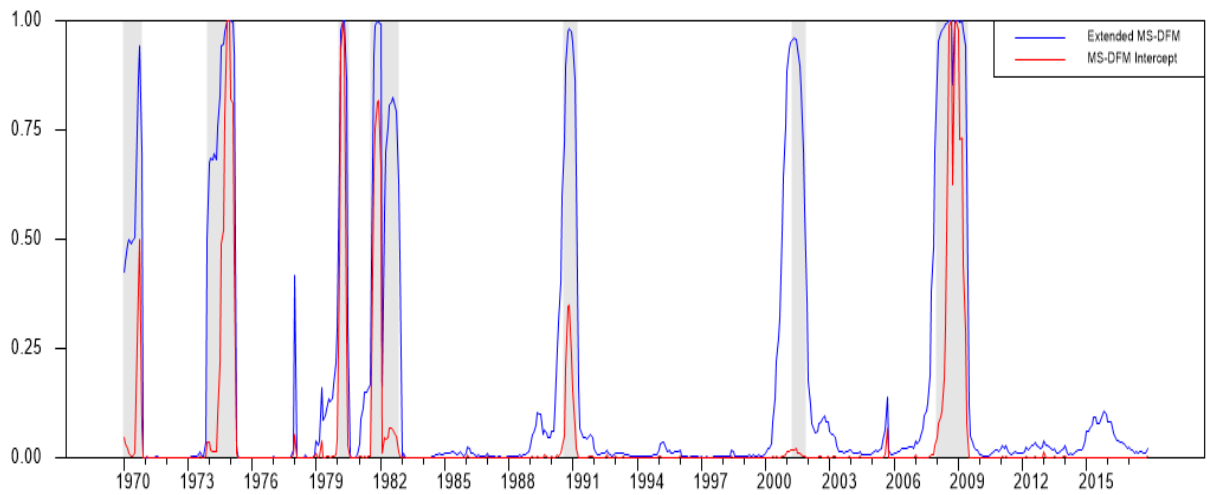
Estimation results are presented in Table 2. Looking at the intercept of the state equation, we notice that the discrepancy between high- and low-volatility recessions is twice bigger than between high-and low-volatility expansions. This asymmetry in the estimated impact of volatility on economic growth is consistent with the vulnerable growth dynamics recently put forward by Adrian *et al.* (2019). These authors show that deteriorating financial conditions, as measured by the National Financial Conditions Index (NFCI), are associated with both an increase in the conditional volatility and a decrease in the conditional mean of quarterly GDP growth in the US. They also show that this effect is asymmetric: the lower quantiles of the distribution of future GDP growth are strongly dependent on the tightness of financial conditions, whereas upper quantiles are stable over time. Looking at historical time series, it appears that financial conditions are precisely tighter than average before the Great Moderation and during the Great Recession (see Figure 2 in Adrian *et al.*, 2019), i.e. when our model identifies a high-volatility regime.

Table 2. Intercept of the state equation, depending on the underlying regime

	Intercept and 95% probability interval	Implied expectation, given $\mathcal{F}1$
High-volatility recession	- 0.78 [- 1.08 ; - 0.45]	- 0.81
Low-volatility recession	- 0.25 [- 0.38 ; - 0.07]	- 0.26
Low-volatility expansion	0.03 [0.00 ; 0.06]	0.03
High-volatility expansion	0.23 [0.07 ; 0.38]	0.23

Note: *Estimation sample: 1970M01-2017M12. Data vintage: 2017M12.*

Figure 2. Influence of switching volatility on smoothed recession probabilities

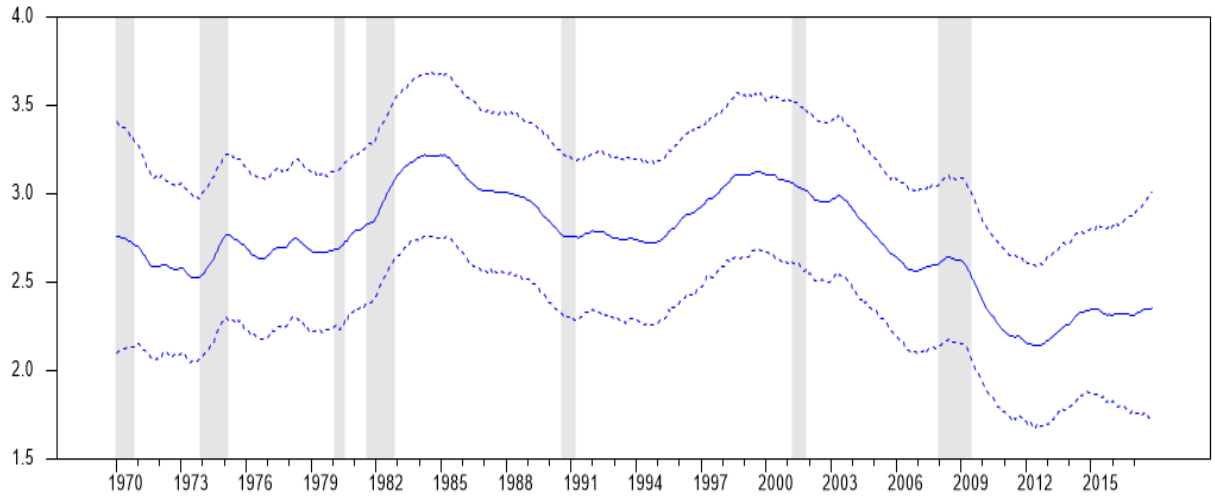


Estimation sample: 1970M01-2017M12. Data vintage: 2017M12. Estimation based on 15 000 draws of the Gibbs Sampler (the first 5 000 are discarded).
MS-DFM Intercept refers to a model where volatility is not allowed to switch. Otherwise, this model has the same characteristics as the Extended MS-DFM.

5.3. Declining long-term GDP growth in the US since the turn of the century

Similarly to Antolin-Diaz *et al.* (2017), we identify a decrease in the US long-term GDP growth which started in the early 2000s, i.e. before the Great Recession, and continued thereafter (see Figure 3). This result is compatible with the view that the Great Recession may have negatively affected the US trend growth rate and reinforced pre-existing factors. As compared to 2000, the estimated US long-term GDP growth is lower by around one percentage point per year in 2010 and seems to have stabilised since then. This slowdown in US long-term GDP growth follows an increase in the second half of the 1990s, which corresponds to a period of increase in US productivity growth (Jorgenson *et al.* 2008).

Figure 3. Extended MS-DFM: Smoothed estimate of the US annual long-run GDP growth rate



Estimation sample: 1970M01-2017M12. Data vintage: 2017M12. Estimation based on 15 000 draws of the Gibbs Sampler (the first 5 000 are discarded).
Dashed lines delineate the 68% credibility interval.

5.4. Comparison of different model specifications based on DICs and marginal likelihoods

The two refinements to the standard MS-DFM framework that we put forward in this paper are not only supported by the fact that they allow to capture changes in long-term GDP growth and to improve turning point detection. They are also supported by statistical criteria.

We first rely on the Deviance Information Criterion (DIC) to compare model specifications. This criterion has been specifically designed to compare complex hierarchical models like the one in this paper (Spiegelhalter *et al.*, 2002). It is also used by Eo and Kim (2016). Like other information criteria, the DIC includes two terms, one capturing model fit and the other penalising for model complexity:

$$DIC = \underbrace{E_{\tilde{\theta},\tilde{S},\tilde{V}|Y}\left[-2\cdot\log f\left(\widetilde{\Delta y_{1..T}}|\tilde{\theta},\widetilde{S_{1..T}},\widetilde{V_{1..T}}\right)\right]}_{\substack{Decreases\ when \\ model\ fit\ improves}} + \underbrace{E_{\tilde{\theta},\tilde{S},\tilde{V}|Y}\left[-2\cdot\log f\left(\widetilde{\Delta y_{1..T}}|\tilde{\theta},\widetilde{S_{1..T}},\widetilde{V_{1..T}}\right)\right] + 2\log f\left(\widetilde{\Delta y_{1..T}}|\tilde{\theta}^*,\widetilde{S_{1..T}}^*,\widetilde{V_{1..T}}^*\right)}_{Penalty\ for\ model\ complexity}$$

In the above formula, $\tilde{\theta}$ corresponds to the vector of constant model parameters (e.g. autoregressive parameters), $\tilde{\theta}^*$ to their posterior mean, and $f\left(\widetilde{\Delta y_{1..T}}|\tilde{\theta},\widetilde{S_{1..T}},\widetilde{V_{1..T}}\right)$ to the model conditional likelihood. Model selection is based on the minimisation of the DIC. This information criterion is relatively easy to compute when model estimation is based on Bayesian Markov Chain Monte Carlo (MCMC) techniques, like Gibbs Sampling. Nevertheless, the literature shows that DIC estimation is all the more accurate that the number of parameters in the conditioning set of the log-likelihood is limited (Chan and Grant 2016). Here, we compute the conditional log-likelihood from the Kalman filter step of the Gibbs Sampler. In a nutshell, the Kalman filter estimates the expectation and variance of the model's state vector, conditional on past observations and other model variables and parameters, which then allows to estimate $\log f\left(\Delta y_t|\widetilde{\Delta y_{1..t-1}},\tilde{\theta},\widetilde{S_{1..T}},\widetilde{V_{1..T}}\right)$ for each date t.[7] In this case, the conditioning set includes the two Markov variables $\widetilde{S_{1..T}}$ and $\widetilde{V_{1..T}}$ and all constant model parameters, but not the large state vector $\widetilde{\alpha_{1..T}}$. Summing these log-predictive likelihoods over time and averaging over posterior draws of the Gibbs Sampler finally gives an estimate of $E_{\tilde{\theta},\tilde{S},\tilde{V}|Y}\left[-2\cdot\log f\left(\widetilde{\Delta y_{1..T}}|\tilde{\theta},\widetilde{S_{1..T}},\widetilde{V_{1..T}}\right)\right]$. Alternatively, first averaging over posterior draws of the Gibbs Sampler, and then computing the predictive log-likelihoods with the Kalman filter allows to estimate $\log f\left(\widetilde{\Delta y_{1..T}}|\tilde{\theta}^*,\widetilde{S_{1..T}}^*,\widetilde{V_{1..T}}^*\right)$.

As an alternative to DICs which might be sensitive to the fact that the high-dimensional Markov variables are included in the conditioning set, we also compute marginal log-likelihoods for different model specifications. In this case, the preferred specification is the one with the highest marginal log-likelihood. As we did for DIC computations, we take advantage of the fact that, conditional on the two Markov variables, the MS-DFM is a linear Gaussian state-space model. Therefore, we can use the Kalman filter to compute $f\left(\Delta y_t|\widetilde{\Delta y_{1..t-1}},\tilde{\theta}^*,\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)$ for each date. The most difficult and time-consuming part of the marginal log-likelihood computation for this class of models is to numerically integrate out $\widetilde{S_{1..t}}$ and $\widetilde{V_{1..t}}$ from $f\left(\Delta y_t|\widetilde{\Delta y_{1..t-1}},\tilde{\theta}^*,\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)$. To do so, we rely on a particle filter which provides us with a sample of N independent particles $\widetilde{S_{1..T}}^n$ and $\widetilde{V_{1..T}}^n$ $(n = 1..N)$. The underlying idea of the particle filter is to draw $\widetilde{S_{1..t}}^n$ and $\widetilde{V_{1..t}}^n$, conditional on $\widetilde{\Delta y_{1..t-1}}$ and $\tilde{\theta}^*$, using an importance sampling distribution which approximates the true but unknown distribution of these variables. An important aspect of the particle filter is that it draws $\widetilde{S_{1..t}}^n$ and $\widetilde{V_{1..t}}^n$ recursively, based on $\widetilde{S_{1..t-1}}^n$ and $\widetilde{V_{1..t-1}}^n$. This allows key computing efficiency gains. The general functioning of the particle filter and key issues such as the choice of the importance sampling distribution and the weighting of the particles are explained in detail in Annex G.

[7] Note that given the structure of the model, the following equality applies $f\left(\Delta y_t|\widetilde{\Delta y_{1..t-1}},\tilde{\theta},\widetilde{S_{1..T}},\widetilde{V_{1..T}}\right) = f\left(\Delta y_t|\widetilde{\Delta y_{1..t-1}},\tilde{\theta},\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)$.

This previous step gives us $f\left(\Delta y_t | \widetilde{\Delta y_{1..t-1}}, \tilde{\theta}^*\right)$ for each date, from which we can compute $\log f\left(\widetilde{\Delta y_{1..T}} | \tilde{\theta}^*\right)$. We finally rely on Chib's (1995) formula to estimate the marginal log-likelihood $\log f\left(\widetilde{\Delta y_{1..T}}\right)$ of each specification, as follows:

$$\log f\left(\widetilde{\Delta y_{1..T}}\right) = \log f\left(\widetilde{\Delta y_{1..T}} | \tilde{\theta}^*\right) + \log \varphi\left(\tilde{\theta}^*\right) - \log \varphi\left(\tilde{\theta}^* | \widetilde{\Delta y_{1..T}}\right)$$

As above, $\tilde{\theta}^*$ corresponds to the vector of constant model parameters evaluated at their posterior mean. $\log \varphi\left(\tilde{\theta}^*\right)$ and $\log \varphi\left(\tilde{\theta}^* | \widetilde{\Delta y_{1..T}}\right)$ are the log-prior and posterior densities of these parameters, respectively. Note that the use of a particle filter to reduce the dimension of the conditioning set prior to the use of Chib's (1995) formula is key to ensure accurate results, as shown by Frühwirth-Schnatter and Wagner (2008).[8]

To the best of our knowledge, Kaufmann (2000) is the only other paper in the literature computing the marginal likelihood of MS-DFMs. However our framework is more complex than the framework of Kaufmann (2000) since we have included Markov-Switching volatility and time-varying growth rates. In this respect, the following results and the marginal likelihood computation method described in Annex G can be considered as additional contributions of the present paper.

[8] An alternative to the use of the particle filter would be to apply Chib's (1995) formula on $\log f\left(\widetilde{\Delta y_{1..T}} | \tilde{\theta}^*, \widetilde{S_{1..T}}^*, \widetilde{V_{1..T}}^*\right)$ directly. This would result in a 'complete-data' marginal likelihood estimator, which would be easier to compute than our proposed estimator. Nevertheless, Frühwirth-Schnatter and Wagner (2008) show that the complete-data marginal likelihood estimator can be extremely inaccurate and recommend not to use it.

Table 3 compares the DICs and marginal log-likelihoods of four model specifications: a linear DFM, a MS-DFM with Markov-Switching in the intercept of the state equation governing factor dynamics, a MS-DFM with Markov-Switching on the intercept and the volatility of shocks of this state equation, and the Extended MS-DFM with the two Markov-Switching features and a time-varying long-term GDP growth rate. All models have the same five observable variables as inputs (see Section 4). This Table shows that both Markov-Switching features reduce the DIC as compared to a linear DFM. This is especially true for the Markov-Switching volatility of shocks. However, allowing for time-variation in the long-term GDP growth rate does not further reduce the DIC. When it comes to the marginal log-likelihood, both Markov-Switching features also improve (i.e. increase) it and we note a further marginal improvement related to the inclusion of the time-varying long-term GDP growth rate into the model. Given the closeness of the DICs and marginal log-likelihoods for the last two models, it seems wise to consider that they cannot be distinguished with these statistical criteria.

Given that GDP is only one out of five observed variables, and that it is only observed once every three months, this could explain why the inclusion of a time-varying trend in the corresponding measurement equation is only marginally supported by the marginal log-likelihood. Focusing on the forecasting performance of GDP after the Great Recession will add further support to this feature (see Section 6).

Table 3. Comparison of model specifications based on DICs and marginal log-likelihoods

Model specification	DIC	Marginal log-likelihood
Linear DFM	3 784.5	-1 929.4
MS-DFM with MS on intercept only	3 725.2	-1 910.8
MS-DFM with MS on intercept and volatility	**3 612.2**	-1 895.4
MS-DFM with MS on intercept and volatility, and time-varying long-term GDP growth rate (i.e. extended MS-DFM)	3 631.7	**-1 892.5**

Note: In all cases, the posterior mean of model parameters is estimated based on 15 000 draws of the Gibbs Sampler, from which the first 5 000 are discarded. In addition, 500 particles are used for marginal log-likelihood computations (see Annex G for details). Sample: 1970M01-2017M12. Data vintage: 2017M12. The preferred specification is the one minimising the DIC and maximising the marginal log-likelihood. In each case, this specification is indicated in bold.

6. Real-time performance of the Extended MS-DFM specification over the last decade (2007-2017)

In this section, we carry out a real-time assessment of our Extended MS-DFM using vintages of data available from January 2007 to January 2017. In this real-time analysis, we start the estimation by considering as a first sample January 1970 - December 2006, then we recursively add new data and re-estimate the model. However, the specification of the model remains unchanged.

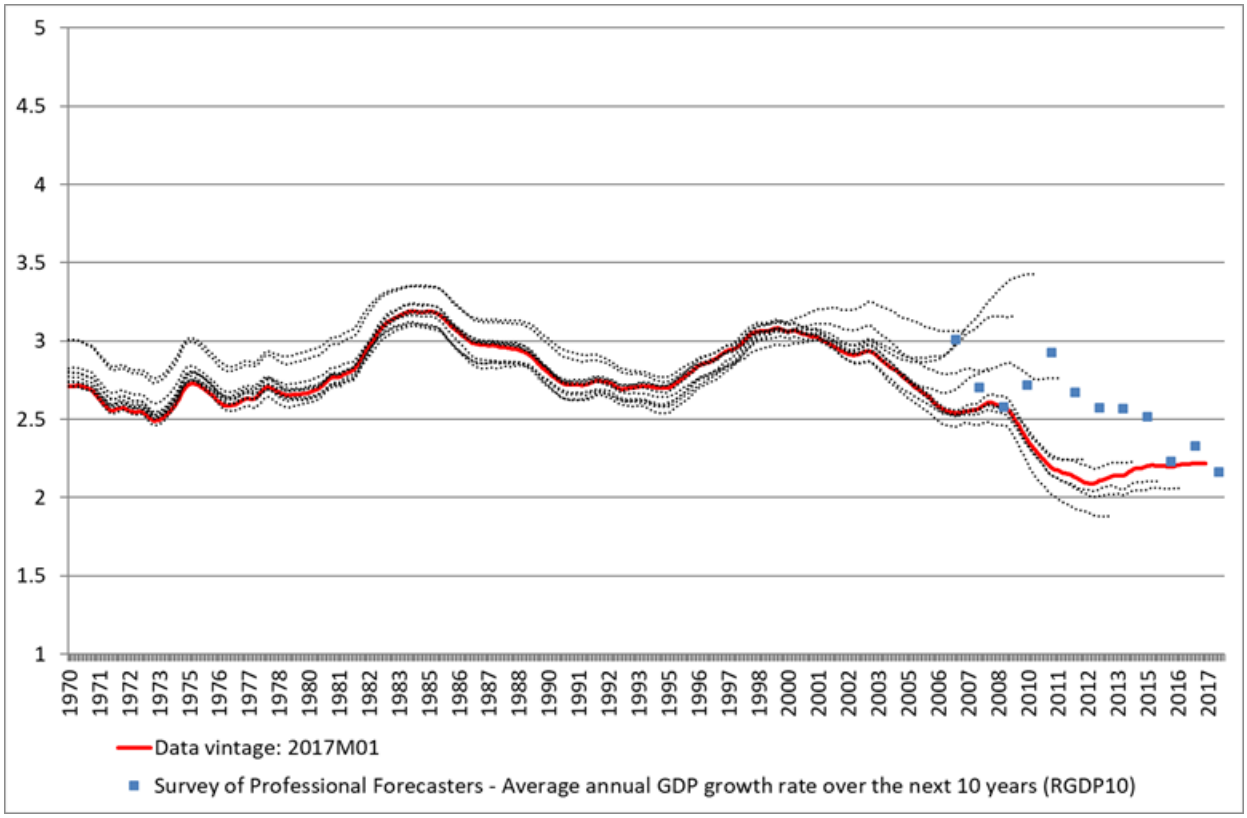
6.1. Real-time estimation of the US long-term GDP growth rate

As already pointed out in the previous section, the decrease in the US long-term GDP growth rate that we estimate with our Extended model is about one percentage point per year compared to the early 2000s (see red curve in Figure 4, corresponding to the 2017M01 data vintage). It is interesting to notice that, over the same period, the professional forecasters surveyed by the Federal Reserve Bank of Philadelphia have revised their expectations of average GDP growth over the next 10 years by a similar amount, and that both estimates converge at the end of the sample.[9]

Figure 4 shows that the model needs some time to identify the drop in long-run GDP growth. Indeed, until 2010, the estimated long-run GDP growth is above 3% but progressively decreases from 2011 onwards (dashed curves in Figure 4). However, when compared with estimates provided by professional forecasters, it appears that the model-based real-time estimates converge more rapidly to the final estimate (red curve) than the SPF real-time estimates since 2011, which seem to be affected by some rigidities.

[9] We use the variable RGDP10 in the Survey of Professional Forecasters (SPF). The corresponding question is asked only once a year, in January. This is why we compare the SPF with our model-based assessment using January data vintages.

Figure 4. Smoothed estimates of the US annual long-term GDP growth rate, consecutive data vintages (2007M01 - 2017M01)



6.2. Real-time detection of the Great Recession

Whereas we relied on the 2017M12 data vintage in order to estimate recession probabilities in Section 5, we now rely on real-time data vintages in order to assess when the model detects the start and the end of the Great Recession.

There is no perfect consensus in the literature on when to formally announce that a recession has started or ended based on the recession probabilities estimated with Markov-Switching models. In the present paper where we use a mix of quarterly and monthly series, we rely on a decision rule which is very similar to the ones advocated by Chauvet and Piger (2008) and Hamilton (2011).[10] We announce a recession when the probability of recession moves from below to above 0.7 and stays above 0.7 for three consecutive months, and we identify as the start of the recession the first month for which the probability of recession exceeds 0.5. Conversely, we announce that a recession is over when the probability of recession moves from above to below 0.3 and stays below 0.3 for

[10] Hamilton (2011) suggests the following decision rule when relying on quarterly GDP data: "When the one-quarter smoothed inference $P(S_t|y_{t+1}, y_t, .., y_1; \hat{\theta}_{t+1})$ first exceeds 0.65, declare that a recession has started, and at that time assign a probable starting point for the recession as the beginning of the most recent set of observations for which $P(S_{t-j}|y_{t+1}, y_t, .., y_1; \hat{\theta}_{t+1})$ exceeds 0.5." Using only the four NBER monthly series to infer recession probabilities, Chauvet and Piger (2008) require that the probability of recession moves from below to above 0.8 and remains above 0.8 for three consecutive months before announcing a recession. Similarly to Hamilton (2011), they then identify as the start of the recession the first month for which the probability moves above 0.5.

three consecutive months, and we identify as the end of the recession the last month for which the probability is above 0.5.

Results are reported in Table 4. The start and end dates of the Great Recession identified by the Extended MS-DFM are in close agreement with those published by the NBER Business Cycle Dating Committee. Nevertheless, the model is able to announce the start and the end of the recession 2 and 11 months earlier than the NBER, respectively, thus allowing to integrate this information into GDP forecasts in a more timely way.

Table 4. Real-time dating of the Great Recession by the NBER and the Extended MS-DFM

Peak date - NBER	Peak date - Extended MS-DFM	Peak date announcement - NBER	Peak date announcement – Extended MS-DFM
2007 M12 (Start of the Great Recession)	2007 M11	2008 M12	2008 M10 (-2m)
Trough date - NBER	**Trough date - Extended MS-DFM**	**Trough date announcement - NBER**	**Trough date announcement – Extended MS-DFM**
2009 M06 (End of the Great Recession)	2009 M06	2010 M09	2009 M10 (-11m)

6.3. Real-time forecasting performance of US quarterly GDP growth

We now focus on real-time GDP forecasting performance in order to highlight the relevance of the Extended MS-DFM specification, with Markov-Switching on the intercept and on the volatility of shocks in the equation governing factor dynamics, as well as a time-varying long-term GDP growth rate. The real-time forecasting performance of this specification is compared to the one of a fully linear DFM specification. The set of observed variables is the same in both cases.

GDP forecasts for the current and the next quarters are produced at the end of each month, from six months up to one month before the quarterly GDP release date by the US Bureau of Economic Analysis (BEA). Given that quarterly GDP for quarter Q is released at the end of the first month of quarter (Q+1), 6-month horizon forecasts are produced at the end of the first month of quarter (Q-1), see Figure 5.

In practice, GDP forecasts are produced based on the first measurement equation of the state-space model. Supplementing the information set with missing data points up to the last month of the next quarter allows the simulation smoother to generate draws of the state vector up to this point, which in turn are used to compute forecasts of quarterly GDP growth. A sequence of forecasts is obtained for each iteration of the Gibbs sampler. These sequences of forecasts are finally averaged across draws to produce the final result.

Figure 5. Calendar of the forecasting exercise

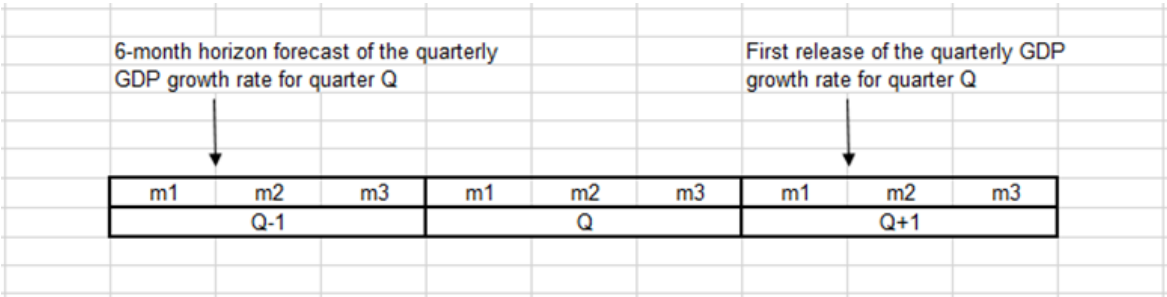


Table 5 compares the real-time Root Mean Square Forecast Errors (RMSFEs) of our Extended MS-DFM and those stemming from a standard linear DFM using the same information set. We present the ratios of RMSFEs: a value below one means that the Extended MS-DFM outperforms the linear DFM. We focus on the Great Recession (2007Q4-2009Q2) and its aftermath (2011Q1-2017Q4), namely the two periods for which we expect regime switches and the time-varying long-term GDP growth rate to have the largest impact on the forecasting performance of the Extended MS-DFM. Figure 6 and Figure 7 compare the underlying point forecasts of the two models at 6-month and 1-month horizons, respectively. These forecasts underlie the RMSFEs shown in Table 5. In terms of RMSFEs, the Extended MS-DFM improves upon the linear DFM at horizons (M-5) and (M-6) in both sub-periods. At shorter horizons, the Extended MS-DFM only remains better than the linear DFM over 2011Q1-2017Q4, but not during the Great Recession.

The better forecasting performance of the Extended MS-DFM as compared to the linear DFM over 2011Q1-2017Q4 is related to the inclusion of the time-varying long-term GDP growth rate into the model. The fact that this feature has such an influence on forecasting performance suggests that the model tends to revert quite quickly to its unconditional mean. As announced in Section 5, focusing on GDP forecasting demonstrates the relevance of the time-varying long-term GDP growth rate even though the impact of this feature is less visible in the DIC and the marginal log-likelihood of the model. We explain this finding by the fact that this feature only plays a role for GDP in the model specification, whereas the DIC and the marginal log-likelihood take into account how the model fits the data for all five variables in the information set simultaneously.

This exercise also shows that the explicit identification of recession regimes helps to improve GDP forecasts six months before the GDP release date. Figure 6 makes it clear that this improvement could be even more substantial if the model were able to identify the start and the end of the Great Recession at earlier dates in real time. Note that the model information set only includes coincident and lagging variables at this stage (Section 4). At least in principle, it should be possible to further improve the detection of transitions between expansion and recession phases by adding leading variables in the model information set. We leave this as an avenue for further research.

Table 5. Relative RMSFEs of the Extended MS-DFM and the linear DFM (forecasts based on real-time data)

	(M-1)	(M-2)	(M-3)	(M-4)	(M-5)	(M-6)
2007Q4-2009Q2	1.02	1.16	1.06	1.07	0.97	0.82
2011Q1-2017Q4	0.98 *(0.22)*	0.95* *(0.07)*	0.85** *(0.02)*	0.92 *(0.11)*	0.88** *(0.04)*	0.90* *(0.06)*

Note: For the comparison over 2011Q1-2017Q4, values in brackets indicate p-values that the RMSFEs of the Extended MS-DFM and the linear DFM are drawn from the same data generating process (H0). Values below 0.1 indicate that H0 can be rejected at the 10% confidence level. These p-values are computed using the Harvey *et al.* (1997) small-sample correction of the Diebold and Mariano (1995) test statistic. As argued by Antolin-Diaz *et al.* (2017) based on Monte Carlo results in Clark and McCracken (2013), this pragmatic approach allows to reliably compare forecasts in cases where models are nested and forecasts are produced using expanding windows. The Great Recession sample (2007Q4-2009Q2) is too short to assess whether differences in forecasting performance between the two models are statistically significant or not.

Figure 6. Comparison of real-time point forecasts at a 6-month horizon: Linear DFM vs. Extended MS-DFM specification

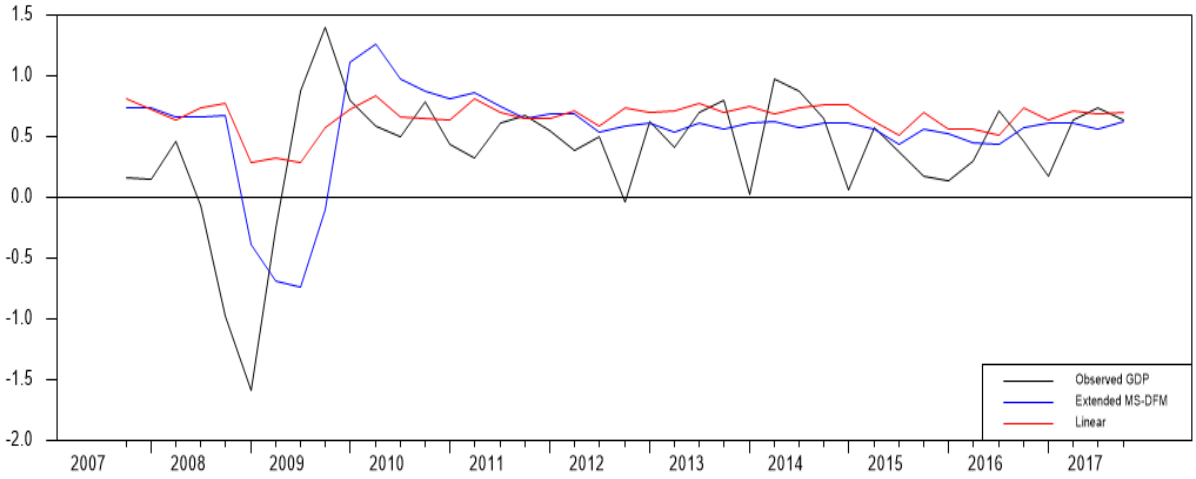
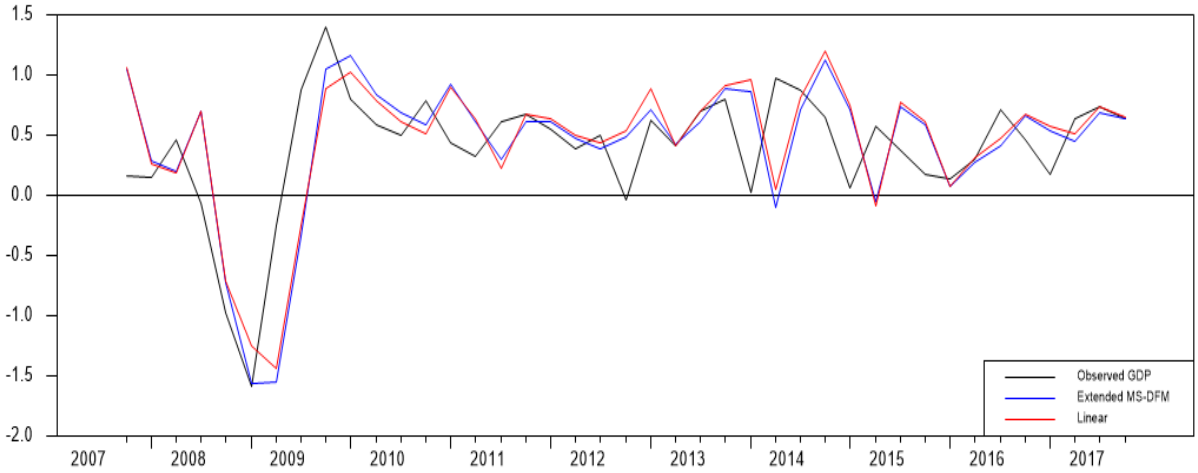


Figure 7. Comparison of real-time point forecasts at a 1-month horizon: linear DFM vs. Extended MS-DFM specification



Finally, Figure 8 and Figure 9 show density forecasts produced by both models at horizons of 6 months. Note that the Bayesian estimation strategy adopted in this paper allows to incorporate all sources of uncertainty in these density forecasts, those related to the filtering of the underlying factor and of the business cycle and volatility states, as well as parameter uncertainties. As opposed to the linear DFM, whose average forecast and forecast uncertainty at a 6-month horizon remain roughly constant over time (Figure 9), the Extended MS-DFM produces a lower average forecast with a larger forecast uncertainty at the time of the Great Recession (Figure 8). It is mainly the left-hand side of the forecast distribution that shifts downwards in this case: low-growth outcomes become more likely when the model identifies a risk of recession, but the upper part of the forecast density remains roughly unchanged. This is in agreement with the vulnerable growth dynamics described by Adrian *et al.* (2019).

Figure 8. Real-time density forecasts at a 6-month horizon: Extended MS-DFM

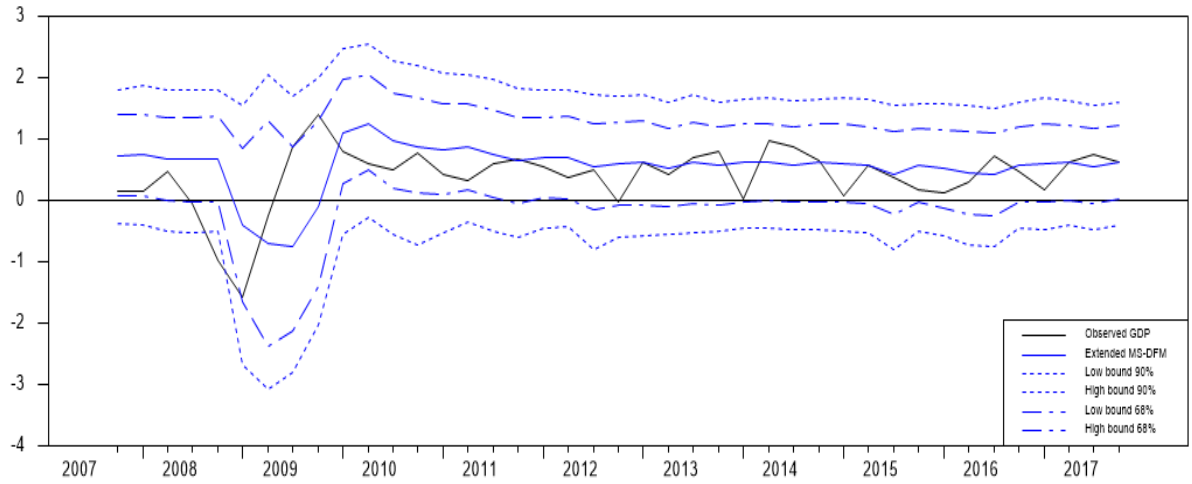


Figure 9. Real-time density forecasts at a 6-month horizon: Linear DFM

7. Conclusion

In this paper, we put forward an extension of the standard Markov-Switching Dynamic Factor Model (MS-DFM) by accounting for two stylised facts of the US economy: changing macroeconomic volatility and fluctuations in the long-run GDP growth rate. This Extended MS-DFM comes as an innovation in the literature and could potentially be used for other advanced economies showing similar patterns.

We first show that the introduction of Markov-Switching volatility in the standard MS-DFM framework is supported by statistical criteria (DICs and marginal likelihoods), and largely improves the detection of turning points in the US business cycle during the Great Moderation. In addition, our Extended model shows evidence that the Great Moderation that started in the mid-1980s is not over, and that the Great Recession, beyond its effect on GDP growth, led to a temporary increase in macroeconomic volatility. The ability of the model to detect recessions in both high- and low-volatility environments is thus very relevant for the detection of future recessions.

As a second improvement to the standard MS-DFM framework, our model also incorporates a time-varying long-term GDP growth rate. This feature enables to identify a continuous decline in the US long-term GDP growth rate of around one percentage point per year as compared to the early 2000s. This result is in line with the empirical literature on the long-run decline in US potential growth. According to our model, about half of this loss occurred before the Great Recession. Nevertheless, the Great Recession may also have negatively affected the US trend growth rate and reinforced pre-existing factors.

Overall, our Extended MS-DFM demonstrates that allowing for both regime switches and time-variation in long-term GDP growth is important to improve the accuracy of real-time short-term GDP forecasts since 2007. The first feature helps to improve GDP forecasts during the Great Recession while the second is most helpful in the aftermath of this recession. Being able to identify the start and the end of the Great Recession even earlier, *e.g.* through the use of leading variables, would allow to further improve the forecasting performance of the model during this and other recessions. We leave this as an avenue for further research.

References

Adrian, T., N. Boyarchenko and D. Giannone (2019), "Vulnerable growth", *American Economic Review*, Vol. 109, No. 4, pp. 1263-1289.

An, Z., J.T. Jalles and P. Loungani (2018), "How well do economists forecast recessions?", *IMF Working Paper,* No. 18/39.

Antolin-Diaz, J., T. Drechsel and I. Petrella (2017), "Tracking the slowdown in long-run GDP growth", *The Review of Economics and Statistics*, Vol. 99, No. 2, pp. 343-356.

Bai, J. and P. Wang (2011), "Conditional Markov chain and its application in economic time series analysis", *Journal of Applied Econometrics*, Vol. 26, pp. 715-734.

Bauwens, L. and J.V.K. Rombouts (2012), "On marginal likelihood computation in change-point models", *Computational Statistics and data Analysis*, Vol. 56, pp. 3415-3429.

Burns, A.F. and W.C. Mitchell (1946), "Measuring business cycles", *National Bureau of Economic Research*.

Camacho, M., G. Perez-Quiros and P. Poncela (2018), "Markov-switching dynamic factor models in real time", International Journal of Forecasting, Vol. 34, pp. 598-611.

Camacho, M., G. Perez-Quiros and P. Poncela (2014), "Green shoots and double dips in the Euro area: A real-time measure", *International Journal of Forecasting*, Vol. 30, pp. 520-535.

Chan, J.C.C. and A.L. Grant (2016), "Fast computation of the deviance information criterion for latent variable models", *Computational Statistics and Data Analysis*, Vol. 100, pp. 847-859.

Charles, A., O. Darné and L. Ferrara (2018), "Does the Great Recession imply the end of the Great Moderation? International evidence", *Economic Inquiry*, Vol. 56, No. 2, pp. 745-760.

Chauvet, M., Z. Senyuz and E. Yoldas (2015), "What does financial volatility tell us about macroeconomic fluctuations?", *Journal of Economic Dynamics and Control*, Vol. 52, pp. 340-360.

Chauvet, M. and Y. Su (2014), "Nonstationarities and Markov switching models", *in Ma and M. Wohar (eds.), Recent Advances in Estimating Nonlinear Models With Applications in Economics and Finance*, Springer Science Business Media, New York.

Chauvet, M. and S. Potter (2013), "Forecasting output", *Handbook of Economic Forecasting*, Chapter 3, pp. 141-194.

Chauvet, M. and J. Piger (2008), "A comparison of the real-time performance of two business-cycle dating methods", *Journal of Business and Economics Statistics*, Vol. 26, No. 1, pp. 42-49.

Chauvet, M. and J.D. Hamilton (2005), "Dating business cycle turning points", *NBER Working Paper*, Vol. 11422.

Chauvet, M. (1998), "An econometric characterization of business cycle dynamics with factor structure and regime switching", *International Economic Review*, Vol. 39, No. 4, pp. 969-996.

Chib, S. (1995), "Marginal likelihood from the Gibbs output", *Journal of the American Statistical Association*, Vol. 90, pp. 1313-1321.

Clark, T.E. and M.W. McCracken (2013), "Advances in forecast evaluation", *Handbook of Economic Forecasting*, Chapter 20, pp. 1107-1201.

Diebold, F.X. and G.D. Rudebusch (1996), "Measuring business cycles: A modern perspective", *The*

Review of Economics and Statistics, Vol. 78, No. 1, pp. 67-77.

Diebold, F.X. and R.S. Mariano (1995), "Comparing predictive accuracy", *Journal of Business and Economic Statistics*, Vol. 13, pp. 253-263.

Doucet, A. and A.M. Johansen (2011), "A tutorial on particle filtering and smoothing: Fifteen years later", *Oxford Handbook of Nonlinear Filtering*, pp. 656-704.

Durbin, J. and S.J. Koopman (2002), "A simple and efficient simulation smoother for state space time series analysis", *Biometrika*, Vol. 89, No. 3, pp. 603-615.

Eo, Y. and J. Morley (2017), "Why has the US economy stagnated since the Great Recession?", *The University of Sydney, Economics Working Paper Series*, Vol. 2017-14.

Eo, Y. and C.-J. Kim (2016), "Markov-switching models with evolving regime-specific parameters: Are postwar booms or recessions all alike?", *The Review of Economics and Statistics*, Vol. 98, No. 5, pp. 940-949.

Ferrara, L. and C. Marsilli (2017), "Global growth: Optimism for 2017?" *EcoNotePad Banque de France, https://blocnotesdeleco.banque-france.fr/en/blog-entry/global-growth-optimism-2017.*

Fernald, J., R. Hall, J. Stock and M. Watson (2017), "The disappointing recovery of output after 2009", *Brooking Papers in Economic Activity*, Vol. 48, No. 1, pp. 1-81.

Frühwirth-Schnatter, S. and H. Wagner (2008), "Marginal likelihoods for non-Gaussian models using auxiliary mixture sampling", *Computational Statistics and Data Analysis*, Vol. 52, pp. 4608-4624.

Gadea-Rivas, M.D., A. Gomez-Loscos and G. Perez-Quiros (2018), "Great Moderation and Great Recession: From plain sailing to stormy seas?" *International Economic Review*, Vol. 59, No. 4, pp. 2297-2321.

Gerlach, R., C. Carter and R. Kohn (2000), "Efficient Bayesian inference for dynamic mixture models", *Journal of the American Statistical Association*, Vol. 95, pp. 819-828.

Giordani, P., R. Kohn and D. van Dijk (2007), "A unified approach to nonlinearity, structural change, and outliers", *Journal of Econometrics*, Vol. 137, pp. 112-133.

Hamilton, J. (2011), "Calling recessions in real time", *International Journal of Forecasting*, Vol. 27, pp. 1006-1026.

Hamilton, J. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle", *Econometrica*, Vol. 57, No. 2, 357-384.

Harding, D. and A. Pagan (2006), "Synchronization of cycles", *Journal of Econometrics*, Vol. 132, pp. 59-79.

Harvey, D., S. Leybourne and P. Newbold (1997), "Testing the equality of prediction mean square errors", *International Journal of Forecasting*, Vol. 13, pp. 281-291.

Jorgenson, D.W., M.S. Ho and K.J. Stiroh (2008), "A retrospective look at the US productivity growth resurgence", *Journal of Economic Perspectives*, Vol. 22, No. 1, pp. 3-24.

Kaufmann, S. (2000), "Measuring business cycles with a dynamic Markov-switching factor model: An assessment using Bayesian simulation methods", *Econometrics Journal*, Vol. 3, pp. 39-65.

Kim, C.J. and C.R. Nelson (1998), "Business cycle turning points, a new coincident index and tests of duration dependence based on a dynamic factor model with regime switching", *The Review of Economics and Statistics*, Vol. 80, No. 2, pp. 188-201.

Koopman, S.-J. and J. Durbin (2003), "Filtering and smoothing of state vector for diffuse state-space models", *Journal of Time Series Analysis*, Vol. 24, No. 1, pp. 85-98.

Koopman, S.J. and J. Durbin (2000), "Fast filtering and smoothing for multivariate state space models", *Journal of Time Series Analysis*, Vol. 21, No. 3, pp. 281-296.

Lewis, C. and N. Pain (2015), "Lessons from OECD forecasts during and after the financial crisis", *OECD Journal: Economic Studies*, Vol. 2014, No. 1, pp. 9-39.

Marcellino, M., M. Porqueddu and F. Venditti (2016), "Short-term forecasting with a mixed-frequency dynamic factor model with stochastic volatility", *Journal of Business and Economic Statistics*, Vol. 34, No. 1, pp. 118-127.

Mariano, R. and Y. Murasawa (2003), "A new coincident index of business cycles based on monthly and quarterly series", *Journal of Applied Econometrics*, Vol. 18, pp. 427-443.

McConnell, M.M. and G. Perez-Quiros (2000), "Output fluctuations in the United States: What has changed since the early 1980s?" *American Economic Review*, Vol. 90, No. 5, pp. 1464-1476.

McCracken, M.W. and S. Ng (2016), "FRED-MD: A monthly database for macroeconomic research", *Journal of Business and Economic Statistics*, Vol. 34, No. 4, pp. 574-589.

OECD (2013), "Well-Being and the Global Financial Crisis", *in How's Life? 2013: Measuring Well-Being. OECD Publishing, Paris. https://doi.org/10.1787/how_life-2013-7-en.*

Sensier, M. and D. van Dijk (2004), "Testing for volatility changes in US macroeconomic time series", *The Review of Economics and Statistics*, Vol. 86, No. 3, pp. 833-839.

Spiegelhalter, D.J., N.G. Best, B.P. Carlin and A. van der Linde (2002), "Bayesian measures of model complexity and fit", *Journal of the Royal Statistical Society*, Series B, Vol. 64, pp. 583-639.

Stiglitz, J.E., J-P. Fitoussi and M. Durand (2018), "Beyond GDP: Measuring What Counts for Economic and Social Performance", *OECD Publishing, Paris. https://doi.org/10.1787/9789264307292-en*.

Stock, J.H. and M.W. Watson (2012), "Disentangling the channels of the 2007-2009 recession", *Brookings Papers on Economic Activity*, Spring 2012, pp. 81-156.

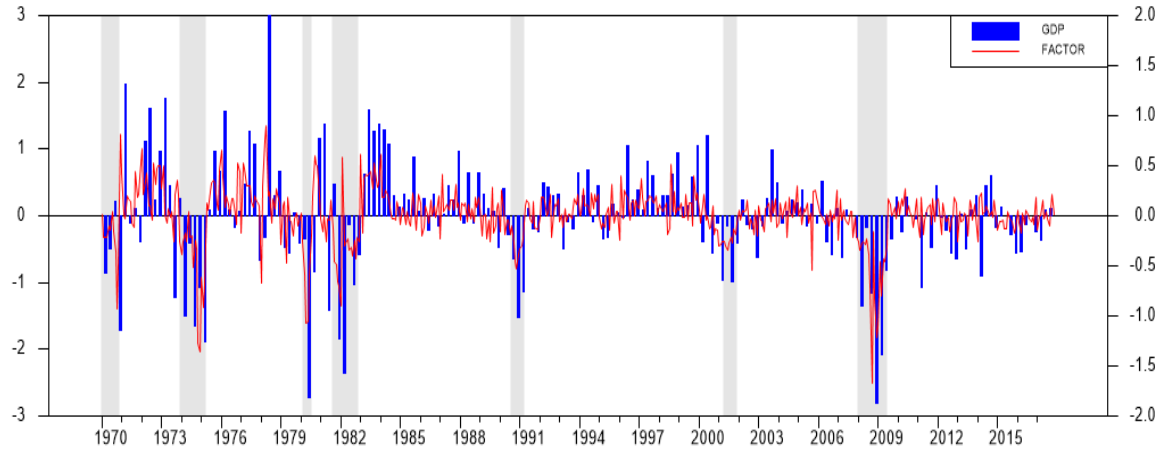
Stock, J.H. and M.W. Watson (2002), "Has the business cycle changed and why?", *NBER Macroeconomics Annual*, MIT Press, pp. 159-230.

Stock, J.H. and M.W. Watson (1989), "New indexes of coincident and leading economic indicators", *NBER Macroeconomics Annual*, MIT Press, pp. 351-394.

Annex A. Input data and underlying factor

In this Annex, we compare the smoothed factor with all input series in the information set, with NBER recession dates and with the smoothed probability of being in a high-volatility regime. All estimates are based on the full model with Markov-Switches in the intercept and the volatility of the state equation (Extended MS-DFM). The estimation sample is 1970M01-2017M12.

Figure A.1. Real GDP (demeaned, blue bars, lhs) and smoothed factor (red, rhs)



Note: Quarterly GDP growth rates are reported in the third month of each quarter

Figure A.2. Industrial production (blue, lhs) and smoothed factor (red, rhs)

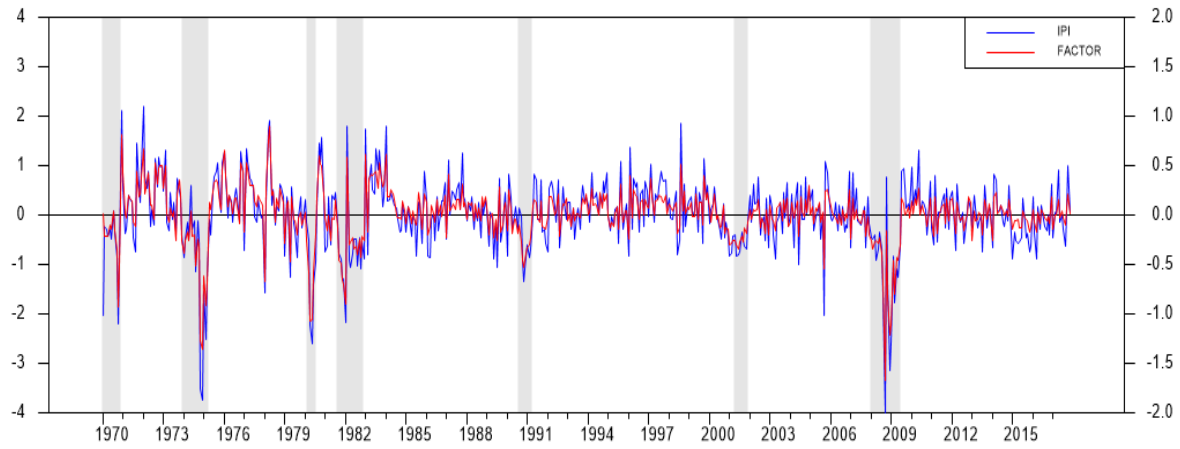


Figure A.3. Real personal income excluding transfer payments (blue, lhs) and smoothed factor (red, rhs)

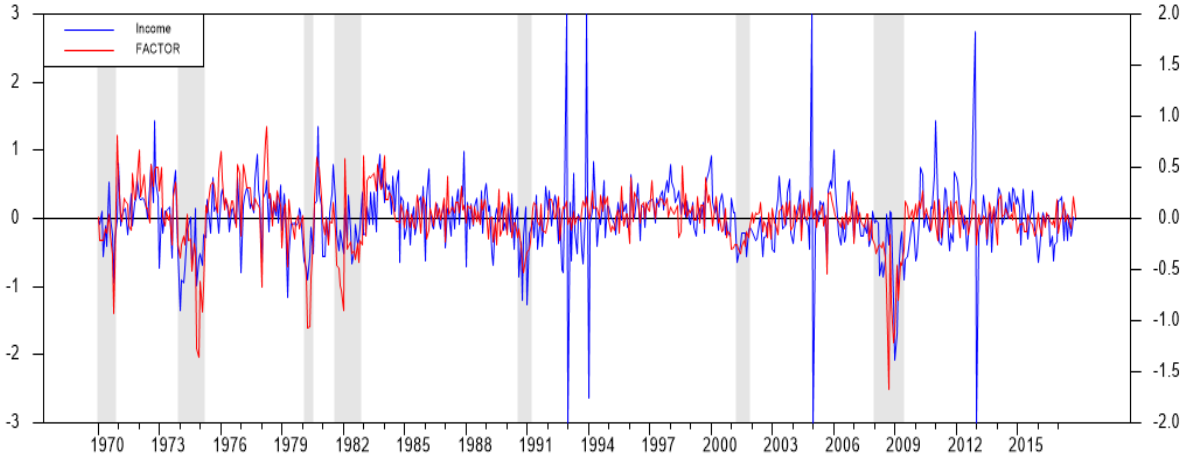


Figure A.4. Real manufacturing trade and sales (blue, lhs) and smoothed factor (red, rhs)

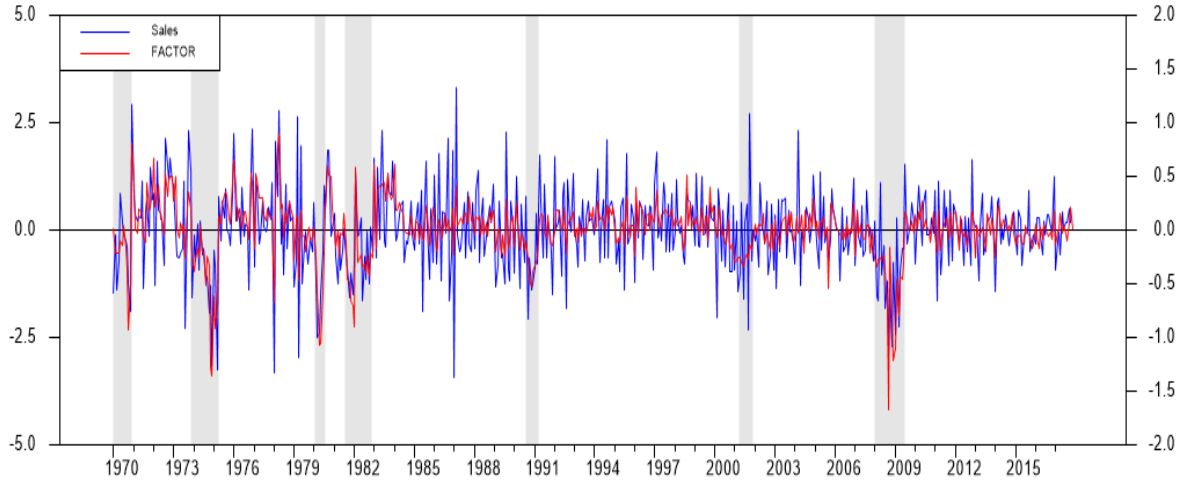


Figure A.5. Non-agricultural civilian employment (blue, lhs) and smoothed factor (red, rhs)

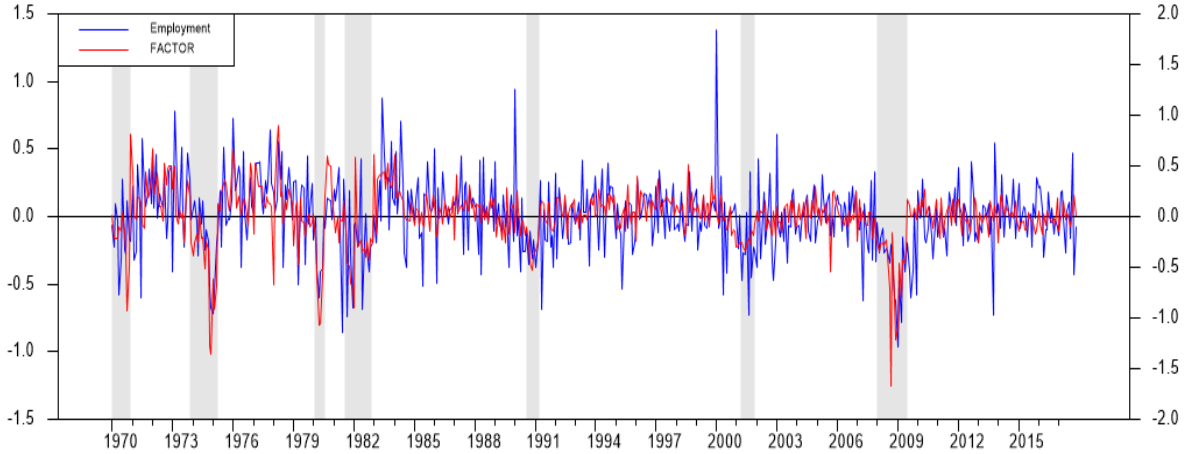
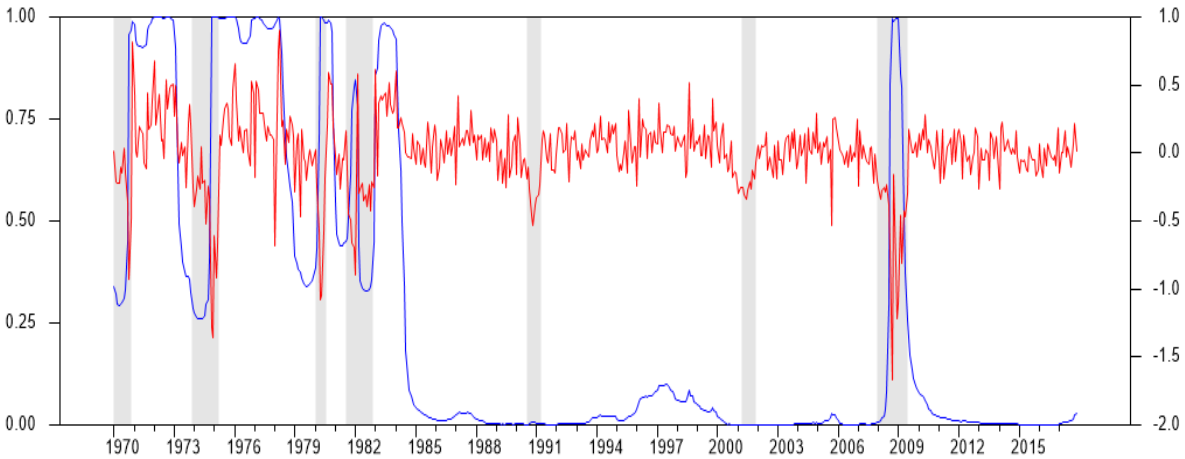


Figure A.6. Underlying factor (red, rhs) and smoothed probability of being in a high-volatility regime (blue, lhs)



Annex B. Influence of the employment variable in the information set

Following Chauvet and Hamilton (2005), we favour non-agricultural civilian employment over non-farm payroll employment[11] in the present paper (Section 4.1). The choice of the employment variable in the information set mainly affects the detection of the 2001 recession and the estimation of the time-varying long-term GDP growth rate. Using non-farm payroll employment significantly delays the dating of the end of the 2001 recession in real time (Table B.1 and Table B.2), as already noted by Chauvet and Hamilton (2005), and implies much larger revisions of the real-time estimates of long-term GDP growth (Figure B.1 and Figure B.2). These two issues are related since failing to identify the end of the 2001 recession forces the model to increase the estimate of long-run GDP growth in order to account for observed GDP growth.

Table B.1. Real-time dating of recessions using non-agricultural civilian employment

Peak date - NBER	Peak date - Extended MS-DFM	Peak date announcement - NBER	Peak date announcement – Extended MS-DFM
2001 M03	2000 M11	2001 M11	2001 M07 (-4m)
2007 M12 (Start of the Great Recession)	2007 M11	2008 M12	2008 M10 (-2m)
Trough date - NBER	**Trough date - Extended MS-DFM**	**Trough date announcement - NBER**	**Trough date announcement – Extended MS-DFM**
2001 M11	2001 M11	2003 M07	2002 M04 (-15m)
2009 M06 (End of the Great Recession)	2009 M06	2010 M09	2009 M10 (-11m)

Table B.2. Real-time dating of recessions using non-farm payroll employment

Peak date - NBER	Peak date - Extended MS-DFM	Peak date announcement - NBER	Peak date announcement – Extended MS-DFM
2001 M03	2000M10	2001 M11	2001 M06 (-5m)
2007 M12 (Start of the Great Recession)	2007 M09	2008 M12	2008 M05 (-7m)
Trough date - NBER	**Trough date - Extended MS-DFM**	**Trough date announcement - NBER**	**Trough date announcement – Extended MS-DFM**
2001 M11	2003 M07	2003 M07	2003 M12 (+5m)
2009 M06 (End of the Great Recession)	2009 M05	2010 M09	2010 M02 (-7m)

[11] In the FRED-MD database, non-agricultural civilian employment and non-farm payroll employment have codes CE16OV and PAYEMS, respectively.

Figure B.1. Smoothed estimates of the US annual long-term GDP growth rate using non-agricultural civilian employment in the information set of the Extended MS-DFM (consecutive data vintages: 2000M01 - 2017M01)

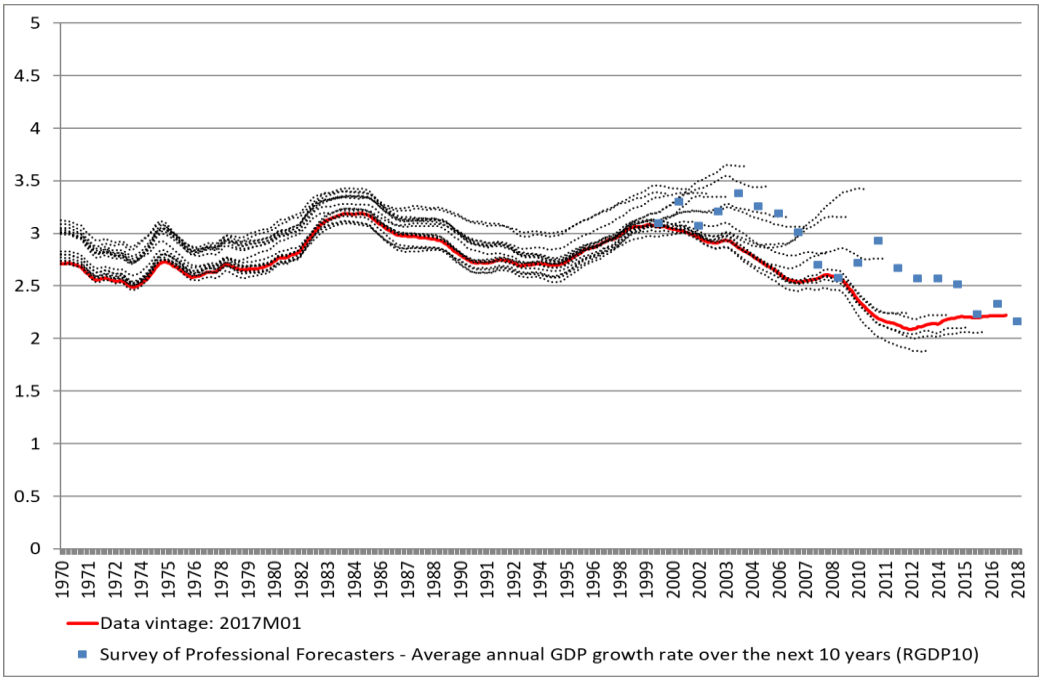
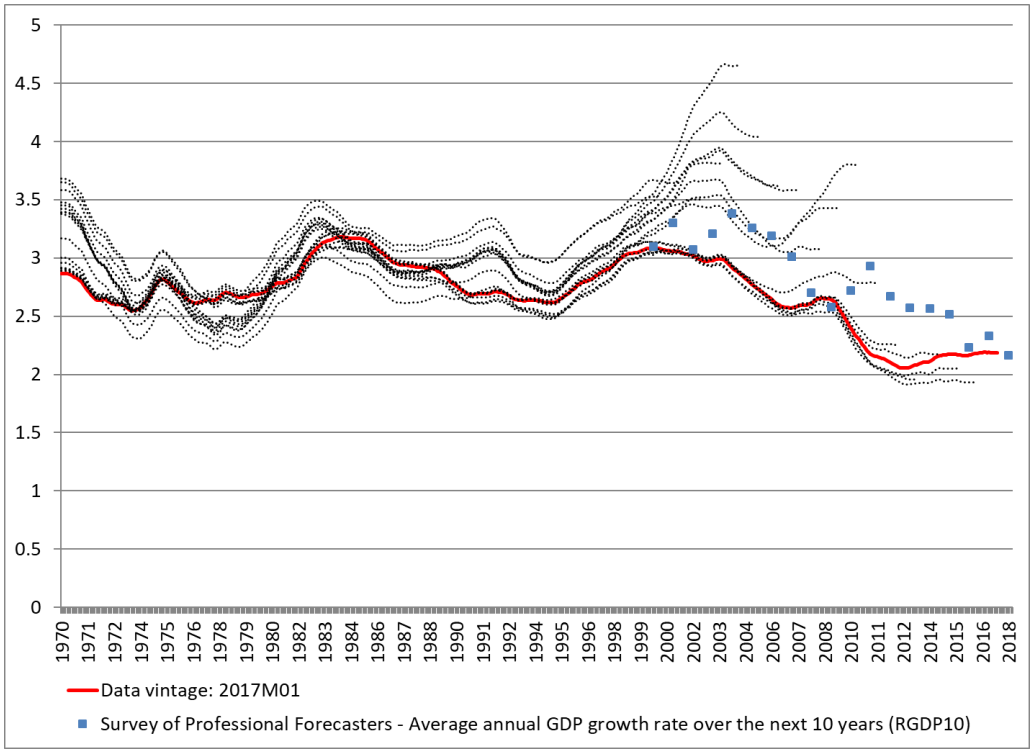


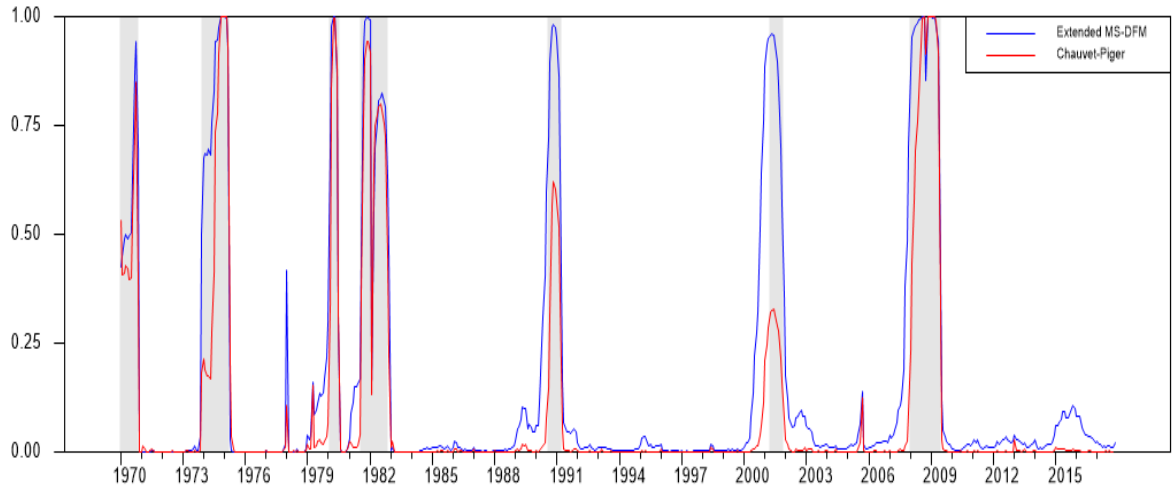
Figure B.2. Smoothed estimates of the US annual long-term GDP growth rate using non-farm payroll employment in the information set of the Extended MS-DFM (consecutive data vintages: 2000M01 - 2017M01)



Annex C. Identification of US business cycle turning points Comparison with Chauvet and Piger (2008)

The comparison with Chauvet and Piger (2008) is based on the updated smoothed recession probabilities that Jeremy Piger publishes every month on its website. [12] We rely on the vintage that he published in December 2017, which is also our last vintage of data. Figure C.1 shows that the Extended MS-DFM specification brings an improvement to the Chauvet-Piger (2008) specification for all recessions. Admittedly, we also get a false but short-lived signal of recession in August 1983.

Figure C.1. Comparison of the smoothed recession probabilities from Chauvet and Piger (2008), updated by J. Piger, and those obtained with the Extended MS-DFM (Dec. 2017)



[12] See http://pages.uoregon.edu/jpiger/us_recession_probs.htm/history_of_real_time_recess.xls.

Annex D. Prior and posterior parameter distributions

Table D.1 gives the first two moments of the prior and posterior parameter distributions, as well as the 95% confidence band of the posterior distributions of the Extended MS-DFM parameters. Note that the prior distributions of all variance parameters are proper, but only their expectations are defined because the first parameter of these inverse-gamma distributions is set equal to two in order to limit the influence of the priors. See Annex E for additional details on the calibration of the priors.

Table D.1. Prior and posterior parameter distributions of the Extended MS-DFM[13]

	Prior			Posterior			
	Distribution	Mean	Standard deviation	Mean	Standard deviation	95% bands	
γ_{10}^{m}	N(0, 1)	0	1	2.07	0.14	1.82	2.37
γ_{20}^{m}	N(0, 1)	0	1	1.90	0.14	1.64	2.19
γ_{30}^{m}	N(0, 1)	0	1	0.77	0.10	0.59	0.96
γ_{40}^{m}	N(0, 1)	0	1	0.28	0.05	0.18	0.38
γ_{41}^{m}	N(0, 1)	0	1	0.09	0.06	-0.02	0.20
γ_{42}^{m}	N(0, 1)	0	1	0.09	0.05	-0.01	0.20
γ_{43}^{m}	N(0, 1)	0	1	0.14	0.05	0.05	0.24
$(\sigma_1^q)^2$	$\mathrm{IG}\left(2, \frac{9}{38} \cdot \mathrm{Var}(\Delta y_{1t}^{q})\right)$	0.15	--	0.24	0.03	0.18	0.31
$(\sigma_1^m)^2$	$\mathrm{IG}(2, \mathrm{Var}(\Delta y_{1t}^{m}))$	0.54	--	0.14	0.02	0.10	0.19
$(\sigma_2^m)^2$	$\mathrm{IG}(2, \mathrm{Var}(\Delta y_{2t}^{m}))$	0.93	--	0.50	0.04	0.43	0.58
$(\sigma_3^m)^2$	$\mathrm{IG}(2, \mathrm{Var}(\Delta y_{3t}^{m}))$	0.38	--	0.32	0.02	0.28	0.36
$(\sigma_4^m)^2$	$\mathrm{IG}(2, \mathrm{Var}(\Delta y_{4t}^{m}))$	0.04	--	0.05	0.003	0.05	0.06
ψ_{11}^{q}	N(0, 1)	0	1	-0.75	0.11	-0.87	-0.45
ψ_{11}^{m}	N(0, 1)	0	1	0.02	0.08	-0.14	0.18
ψ_{21}^{m}	N(0, 1)	0	1	-0.37	0.04	-0.45	-0.29
ψ_{31}^{m}	N(0, 1)	0	1	-0.15	0.04	-0.24	-0.07
ψ_{41}^{m}	N(0, 1)	0	1	-0.16	0.04	-0.25	-0.08
ϕ_1	N(0, 1)	0	1	0.04	0.12	-0.17	0.28
$(\sigma_c)^2$	$\mathrm{IG}\left(2, \frac{9}{38} \cdot \mathrm{Var}(\Delta y_{1t}^{q})\right)$	0.15	--	0.04	0.01	0.02	0.05
$(\sigma_{a,1})^2$	IG(2, 5e-3)	5.00e-3	--	9.24e-5	4.69e-5	5.13e-5	1.30e-4
PrS_{00}	Beta(90, 7)	0.93	2.61e-2	0.91	2.34e-2	0.86	0.95
PrS_{11}	Beta(470, 7)	0.99	5.50e-3	0.98	4.62e-3	0.97	0.99
PrV_{00}	Beta(285, 1)	0.99	2.58e-3	0.99	4.51e-3	0.98	1.00
PrV_{11}	Beta(285, 1)	0.99	2.58e-3	0.98	8.34e-3	0.97	1.00
h	IG(5, 10)	2.50	1.44	0.85	0.53	0.04	1.99
μ_{00}	N(0, 1)	0	1	-0.25	0.08	-0.38	-0.07
μ_{10}	N(0, 1)	0	1	0.28	0.08	0.10	0.42
μ_{01}	N(0, 1)	0	1	-0.54	0.14	-0.77	-0.25
μ_{11}	N(0, 1)	0	1	0.73	0.19	0.36	1.07

[13] Estimation sample: 1970M01-2017M12. Data vintage: 2017M12. Remind that γ_{10}^{q} is set equal to 1 for identification purposes, and not estimated (see Section 2 and Annex E).

Annex E. State-space representation of the model and description of the estimation methodology

State-space representation of the model

For estimation purposes, it is useful to consider two different state-space representations of the MS-DFM with Markov-Switching volatility and time-variation in the long-term GDP growth rate introduced in this paper. Notations are the same as in the main text.

Representation 1:

$$\begin{cases} \Delta y_{1t}^q = \left(\frac{1}{3}a_{1,t}^q + \frac{2}{3}a_{1,t-1}^q + a_{1,t-2}^q + \frac{2}{3}a_{1,t-3}^q + \frac{1}{3}a_{1,t-4}^q\right) + \gamma_{10}^q\left(\frac{1}{3}\Delta c_t + \frac{2}{3}\Delta c_{t-1} + \Delta c_{t-2} + \frac{2}{3}\Delta c_{t-3} + \frac{1}{3}\Delta c_{t-4}\right) \\ \qquad + \left(\frac{1}{3}u_{1,t}^q + \frac{2}{3}u_{1,t-1}^q + u_{1,t-2}^q + \frac{2}{3}u_{1,t-3}^q + \frac{1}{3}u_{1,t-4}^q\right) \\ \Delta y_{jt}^m = \gamma_j^m(L)\Delta c_t + u_{j,t}^m \quad \text{for } j = 1 \dots 4 \\ a_{1,t}^q = a_{1,t-1}^q + \sigma_a \cdot \eta_t^a \quad ; \quad \eta_t^a \sim N(0,1) \\ \phi(L)\Delta c_t = \mu_{S_t,V_t} + \sqrt{1 + h \cdot V_t} \cdot \sigma_c \cdot \eta_t^c \quad ; \quad \eta_t^c \sim N(0,1) \\ \psi_1^q(L)u_{1,t}^q = u_{1,t}^q - \psi_{11}^q u_{1,t-1}^q = \sigma_1^q \cdot \varepsilon_{1t}^q \quad ; \quad \varepsilon_{1t}^q \sim N(0,1) \\ \psi_j^m(L)u_{j,t}^m = u_{j,t}^m - \psi_{11}^m u_{j,t-1}^m = \sigma_j^m \cdot \varepsilon_{jt}^m \quad ; \quad \varepsilon_{jt}^m \sim N(0,1) \quad \text{for } j = 1 \dots 4 \end{cases}$$

Representation 2:

$$\begin{cases} \Delta y_{1t}^q = \left(\frac{1}{3}a_{1,t}^q + \frac{2}{3}a_{1,t-1}^q + a_{1,t-2}^q + \frac{2}{3}a_{1,t-3}^q + \frac{1}{3}a_{1,t-4}^q\right) + \gamma_{10}^q\left(\frac{1}{3}\Delta c_t + \frac{2}{3}\Delta c_{t-1} + \Delta c_{t-2} + \frac{2}{3}\Delta c_{t-3} + \frac{1}{3}\Delta c_{t-4}\right) \\ \qquad + \left(\frac{1}{3}u_{1,t}^q + \frac{2}{3}u_{1,t-1}^q + u_{1,t-2}^q + \frac{2}{3}u_{1,t-3}^q + \frac{1}{3}u_{1,t-4}^q\right) \\ \psi_j^m(L)\Delta y_{jt}^m \equiv \Delta y_{jt}^{m,*} = \underbrace{\gamma_j^m(L)\psi_j^m(L)}_{\equiv \gamma_j^{m,*}(L)}\Delta c_t + \sigma_j^m \cdot \varepsilon_{jt}^m = \gamma_j^m(L)\underbrace{\psi_j^m(L)\Delta c_t}_{\equiv \Delta c_{jt}^*} + \sigma_j^m \cdot \varepsilon_{jt}^m \quad ; \quad \varepsilon_{jt}^m \sim N(0,1) \quad \text{for } j = 1 \dots 4 \\ a_{1,t}^q = a_{1,t-1}^q + \sigma_a \cdot \eta_t^a \quad ; \quad \eta_t^a \sim N(0,1) \\ \phi(L)\Delta c_t = \mu_{S_t,V_t} + \sqrt{1 + h \cdot V_t} \cdot \sigma_c \cdot \eta_t^c \quad ; \quad \eta_t^c \sim N(0,1) \\ \psi_1^q(L)u_{1,t}^q = u_{1,t}^q = \psi_{11}^q u_{1,t-1}^q + \sigma_1^q \cdot \varepsilon_{1t}^q \quad ; \quad \varepsilon_{1t}^q \sim N(0,1) \end{cases}$$

Using representation 2, the model can be cast in state-space from, as follows:

Measurement equation:

$$\underbrace{\begin{pmatrix} \Delta y_{1t}^{q} \\ \Delta y_{t}^{m,*} \end{pmatrix}}_{\equiv \Delta y_t^*} = \underbrace{\begin{pmatrix} \rho & \gamma_{10}^{q}\,\rho & \rho \\ 0_{4,5} & \Gamma^* & 0_{4,5} \end{pmatrix}}_{\equiv Z} \begin{pmatrix} A_t^q \\ \Delta C_t \\ U_t^q \end{pmatrix} + \begin{pmatrix} 0 \\ \Sigma_m\, \varepsilon_t^m \end{pmatrix}$$

where:

$$\Delta y_t^{m,*} = (\Delta y_{1t}^{m,*}, \dots, \Delta y_{4t}^{m,*})'$$

$$A_t^q = \left(a_{1,t}^q, a_{1,t-1}^q, a_{1,t-2}^q, a_{1,t-3}^q, a_{1,t-4}^q\right)'$$

$$\Delta C_t = (\Delta c_t, \Delta c_{t-1}, \Delta c_{t-2}, \Delta c_{t-3}, \Delta c_{t-4})'$$

$$U_t^q = \left(u_{1t}^q, u_{1,t-1}^q, u_{1,t-2}^q, u_{1,t-3}^q, u_{1,t-4}^q\right)'$$

$$\varepsilon_t^m = (\varepsilon_{1t}^m, \dots, \varepsilon_{4t}^m)'$$

$$\Sigma_m = \mathrm{diag}(\sigma_1^m, \dots, \sigma_4^m)$$

$$\rho = \left(\frac{1}{3}, \frac{2}{3}, 1, \frac{2}{3}, \frac{1}{3}\right)$$

$$\Gamma^* = \begin{pmatrix} \gamma_{10}^{m,*} & \gamma_{11}^{m,*} & 0 & 0 & 0 \\ \gamma_{20}^{m,*} & \gamma_{21}^{m,*} & 0 & 0 & 0 \\ \gamma_{30}^{m,*} & \gamma_{31}^{m,*} & 0 & 0 & 0 \\ \gamma_{40}^{m,*} & \gamma_{41}^{m,*} & \gamma_{42}^{m,*} & \gamma_{43}^{m,*} & \gamma_{44}^{m,*} \end{pmatrix}$$

With $\gamma_{j1}^m = \gamma_{j2}^m = \gamma_{j3}^m = 0$ for $j = 1..3$, the following applies:

$$\begin{cases} \gamma_{j0}^{m,*} = \gamma_{j0}^m \\ \gamma_{11}^{m,*} \equiv -\gamma_{10}^m \psi_{11}^m \\ \gamma_{21}^{m,*} \equiv -\gamma_{20}^m \psi_{21}^m \\ \gamma_{31}^{m,*} \equiv -\gamma_{30}^m \psi_{31}^m \\ \gamma_{41}^{m,*} \equiv -\gamma_{40}^m \psi_{41}^m + \gamma_{41}^m \\ \gamma_{42}^{m,*} \equiv -\gamma_{41}^m \psi_{41}^m + \gamma_{42}^m \\ \gamma_{43}^{m,*} \equiv -\gamma_{42}^m \psi_{41}^m + \gamma_{43}^m \\ \gamma_{44}^{m,*} \equiv -\gamma_{43}^m \psi_{41}^m \end{cases}$$

State equation:

$$\underbrace{\begin{pmatrix} A_t^q \\ \Delta C_t \\ U_t^q \end{pmatrix}}_{\equiv \alpha_t} = \begin{pmatrix} 0_{5,1} \\ \mu_{S_t,V_t} \cdot \delta \\ 0_{9,1} \end{pmatrix} + \underbrace{\begin{pmatrix} J & 0_{5,5} & 0_{5,5} \\ 0_{5,5} & \Phi & 0_{5,5} \\ 0_{5,5} & 0_{5,5} & \Psi \end{pmatrix}}_{\equiv T} \underbrace{\begin{pmatrix} A_{t-1}^q \\ \Delta C_{t-1} \\ U_{t-1}^q \end{pmatrix}}_{\equiv \alpha_{t-1}} + \begin{pmatrix} \sigma_a \cdot \eta_t^a \\ \sqrt{1 + h \cdot V_t} \cdot \sigma_c \cdot \eta_t^c \\ \sigma_1^q \cdot \varepsilon_{1t}^q \end{pmatrix} \otimes \delta$$

where:

$$\delta = (1 \quad 0 \quad 0 \quad 0 \quad 0)'$$

$$J = \begin{pmatrix} 1 & 0_{1,4} \\ I_4 & 0_{4,1} \end{pmatrix}$$

$$\Phi = \begin{pmatrix} \phi_1 & 0_{1,4} \\ I_4 & 0_{4,1} \end{pmatrix}$$

$$\Psi = \begin{pmatrix} \psi_{11}^q & 0_{1,4} \\ I_4 & 0_{4,1} \end{pmatrix}$$

Identification of the model

Kim and Nelson (1998) consider a model without volatility switches, nor low-frequency fluctuations:

$$\begin{cases} \Delta y_{it} = \gamma_i(L)\Delta c_t + u_{it} \\ \psi_i(L)u_{it} = \sigma_i \cdot \varepsilon_{it} \quad ; \quad \varepsilon_{it} \sim IID\ N(0,1) \\ \phi(L)\left(\Delta c_t - \mu_{S_t}\right) = \sigma_c \cdot \eta_t^c \quad ; \quad \eta_t^c \sim N(0,1) \end{cases}$$

For identification, they impose $\sigma_c = 1$. In our case, the volatility of shocks in the state equation is left free to vary over time and we achieve identification by imposing $\gamma_{10}^q = 1$.

Estimation methodology

The proposed estimation methodology relies on a Gibbs Sampling algorithm which aims at simulating draws from $p(\widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\gamma}, \tilde{\psi}, \sigma_1^q, \widetilde{\sigma^m}, \phi_1, \tilde{\mu}, h, \sigma_c, \sigma_a, PrS_{00}, PrS_{11}, PrV_{00}, PrV_{11} | \widetilde{\Delta y_{1..T}})$. The algorithm sequentially draws from the following conditional distributions:

$$p(\widetilde{\alpha_{1..T}} | \widetilde{\Delta y_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta})$$
$$p(\widetilde{S_{1..T}} | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}) = p(\widetilde{S_{1..T}} | \widetilde{\Delta y_{1..T}}, \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta})$$
$$p(\widetilde{V_{1..T}} | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \tilde{\theta}) = p(\widetilde{V_{1..T}} | \widetilde{\Delta y_{1..T}}, \widetilde{\Delta c_{1..T}}, \widetilde{S_{1..T}}, \tilde{\theta})$$
$$p(\tilde{\gamma} | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta} \backslash \tilde{\gamma}) = p(\tilde{\gamma} | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \tilde{\psi}, \widetilde{\sigma^m})$$
$$p(\tilde{\psi} | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta} \backslash \tilde{\psi}) = p(\tilde{\psi} | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \tilde{\gamma}, \widetilde{\sigma^m}, \sigma_1^q)$$
$$p(\sigma_1^q, \widetilde{\sigma^m} | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta} \backslash \{\sigma_1^q, \widetilde{\sigma^m}\}) = p(\sigma_1^q, \widetilde{\sigma^m} | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \tilde{\gamma}, \tilde{\psi})$$
$$p(\phi_1 | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta} \backslash \phi_1) = p(\tilde{\phi} | \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\mu}, h, \sigma_c)$$
$$p(\tilde{\mu} | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta} \backslash \tilde{\mu}) = p(\tilde{\mu} | \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \phi_1, h, \sigma_c)$$
$$p(h | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta} \backslash h) = p(h | \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \phi_1, \tilde{\mu}, \sigma_c)$$
$$p(\sigma_c | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta} \backslash \sigma_c^2) = p(\sigma_c | \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \phi_1, \tilde{\mu}, h)$$
$$p(\sigma_a | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta} \backslash \sigma_a) = p(\sigma_a | \widetilde{\alpha_{1..T}})$$
$$p(PrS_{00}, PrS_{11} | \widetilde{\Delta y_{1..T}}, \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta} \backslash \{PrS_{00}, PrS_{11}\}) = p(PrS_{00}, PrS_{11} | \widetilde{S_{1..T}})$$
$$p(PrV_{00}, PrV_{11} | \widetilde{\Delta y_T}, \widetilde{\alpha_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta} \backslash \{PrV_{00}, PrV_{11}\}) = p(PrV_{00}, PrV_{11} | \widetilde{V_{1..T}})$$

where the following notations are used:

$$\widetilde{\Delta y_{1..T}} \equiv (\Delta y_1, \dots, \Delta y_T) \; ; \; \widetilde{\alpha_{1..T}} \equiv (\alpha_1, \dots, \alpha_T) \; ; \; \widetilde{S_{1..T}} \equiv (S_1, \dots, S_T) \; ; \; \widetilde{V_{1..T}} \equiv (V_1, \dots, V_T)$$
$$\tilde{\gamma} \equiv (\gamma_{10}^q, \gamma_{10}^m, \gamma_{20}^m, \gamma_{30}^m, \gamma_{40}^m, \gamma_{41}^m, \gamma_{42}^m, \gamma_{43}^m)$$
$$\tilde{\psi} \equiv (\psi_{11}^q, \psi_{11}^m, \psi_{21}^m, \psi_{31}^m, \psi_{41}^m)$$
$$\widetilde{\sigma^m} \equiv (\sigma_1^m, \sigma_2^m, \sigma_3^m, \sigma_4^m)$$
$$\tilde{\mu} \equiv (\mu_{00}, \mu_{01}, \mu_{10}, \mu_{11})$$
$$\tilde{\theta} \equiv (\tilde{\gamma}, \tilde{\psi}, \sigma_1^q, \widetilde{\sigma^m}, \phi_1, \tilde{\mu}, h, \sigma_c, \sigma_a, PrS_{00}, PrS_{11}, PrV_{00}, PrV_{11})$$

Step 1: Based on Representation 2, **draw** $\widetilde{\boldsymbol{\alpha}_{1..T}}$ conditional on $\widetilde{\Delta y^*_{1..T}}$, $\widetilde{S_{1..T}}$, $\widetilde{V_{1..T}}$ and model parameters, based on the sequential Kalman filter with diffuse initialisation of Koopman and Durbin (2000, 2003) and the simulation smoother of Durbin and Koopman (2002).

Step 2: Based on Representation 2, **draw** $\widetilde{\boldsymbol{S_{1..T}}}$ conditional on $\widetilde{\Delta y^*_{1..T}}$, $\widetilde{\Delta c_{1..T}}$, $\widetilde{V_{1..T}}$ and model parameters:[14]

$$p\left(\widetilde{S_{1..T}}|\widetilde{\Delta y^*_{1..T}},\widetilde{\Delta c_{1..T}},\widetilde{V_{1..T}},\tilde{\theta}\right)=p\left(S_T|\widetilde{\Delta c_{1..T}},\widetilde{V_{1..T}},\tilde{\theta}\right)\cdot\prod_{t=1}^{T-1}p\left(S_t|S_{t+1},\widetilde{\Delta c_{1..t}},\widetilde{V_{1..T}},\tilde{\theta}\right)$$

Step 2a: Run Hamilton's (1989) filter to get $p\left(S_T|\widetilde{\Delta c_{1..T}},\widetilde{V_{1..T}},\tilde{\theta}\right)$ and $p\left(S_t|\widetilde{\Delta c_{1..t}},\widetilde{V_{1..T}},\tilde{\theta}\right)$ for t = 1..T-1

Step 2a1: Initialise the filter with unconditional probabilities

$$p\left(S_0=j|\Delta c_0,\widetilde{V_{1..T}},\tilde{\theta}\right)=p(S_0=j)$$

Step 2a2: from $p\left(S_{t-1}=j|\Delta\widetilde{c_{1..t-1}},\widetilde{V_T},\tilde{\theta}\right)$ to $p\left(S_t=k|\Delta\widetilde{c_{1..t-1}},\widetilde{V_T},\tilde{\theta}\right)$

$$p\left(S_t=k|\Delta\widetilde{c_{1..t-1}},\widetilde{V_{1..T}},\tilde{\theta}\right)=\sum_j p\left(S_t=k,S_{t-1}=j|\Delta\widetilde{c_{1..t-1}},\widetilde{V_{1..T}},\tilde{\theta}\right)$$

$$=\sum_j p\left(S_t=k|S_{t-1}=j,\cancel{\Delta\widetilde{c_{1..t-1}},\widetilde{V_{1..T}},\tilde{\theta}}\right)\cdot p\left(S_{t-1}=j|\Delta\widetilde{c_{1..t-1}},\widetilde{V_{1..T}},\tilde{\theta}\right)$$

Step 2a3: from $p\left(S_t=k|\Delta\widetilde{c_{1..t-1}},\widetilde{V_{1..T}},\tilde{\theta}\right)$ to $p\left(S_t=k|\widetilde{\Delta c_{1..t}},\widetilde{V_{1..T}},\tilde{\theta}\right)$

$$p\left(S_t=k|\widetilde{\Delta c_{1..t}},\widetilde{V_{1..T}},\tilde{\theta}\right)=\frac{f\left(\Delta c_t|S_t=k,\Delta\widetilde{c_{1..t-1}},\widetilde{V_{1..T}},\tilde{\theta}\right)\cdot p\left(S_t=k|\Delta\widetilde{c_{1..t-1}},\widetilde{V_{1..T}},\tilde{\theta}\right)}{f\left(\Delta c_t|\Delta\widetilde{c_{1..t-1}},\widetilde{V_{1..T}},\tilde{\theta}\right)}$$

$$f\left(\Delta c_t|\Delta\widetilde{c_{1..t-1}},\widetilde{V_{1..T}},\tilde{\theta}\right)=\sum_k f\left(\Delta c_t|S_t=k,\Delta\widetilde{c_{1..t-1}},\widetilde{V_T},\tilde{\theta}\right)\cdot p\left(S_t=k|\Delta\widetilde{c_{1..t-1}},\widetilde{V_{1..T}},\tilde{\theta}\right)$$

Step 2a4: Store $\boldsymbol{p\left(S_t=k|\widetilde{\Delta c_{1..t}},\widetilde{V_{1..T}},\tilde{\theta}\right)}$ as an intermediate output of Hamilton's filter.

Restart at Step 2a2 as long as $t\leq T-1$.

[14] See proof in Annex F.

Step 2b: from $p\left(S_t|\widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$ to $p\left(S_t|S_{t+1}, \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$

$$p\left(S_t|S_{t+1}, \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) = \frac{p\left(S_{t+1}|S_t, \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) \cdot p\left(S_t|\widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}{p\left(S_{t+1}|\widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$$

$$= \frac{p(S_{t+1}|S_t) \cdot \boldsymbol{p\left(S_t|\widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}}{\sum_j p\left(S_{t+1}|S_t = j, \cancel{\widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}}\right) \cdot p\left(S_t = j|\widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$$

$\boldsymbol{p\left(S_t|\widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$ is an intermediate output of Hamilton's filter (Step 2a). A random number drawn from a uniform distribution over [0,1] is used to generate S_t based on the above formula.

Step 3: Based on Representation 2, **draw** $\boldsymbol{\widetilde{V_{1..T}}}$ conditional on $\widetilde{\Delta y^*_{1..T}}$, $\widetilde{\Delta c_{1..T}}$, $\widetilde{S_{1..T}}$ and model parameters:[15]

$$p\left(\widetilde{V_{1..T}}|\widetilde{\Delta y^*_{1..T}}, \widetilde{\Delta c_{1..T}}, \widetilde{S_{1..T}}, \tilde{\theta}\right) = p\left(V_T|\widetilde{\Delta c_{1..T}}, \widetilde{S_{1..T}}, \tilde{\theta}\right) \cdot \prod_{t=1}^{T-1} p\left(V_t|V_{t+1}, \widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)$$

Step 3a: Run Hamilton's (1989) filter to get $p\left(V_T|\widetilde{\Delta c_{1..T}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)$ and $p\left(V_t|\widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)$ for t = 1..T-1

Step 3a1: Initialise the filter with unconditional probabilities

$$p\left(V_0 = j|\Delta c_0, \widetilde{S_{1..T}}, \tilde{\theta}\right) = p(V_0 = j)$$

Step 3a2: from $p\left(V_{t-1} = j|\Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)$ to $p\left(V_t = k|\Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)$

$$p\left(V_t = k|\Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}\right) = \sum_j p\left(V_t = k, V_{t-1} = j|\Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)$$

$$= \sum_j p\left(V_t = k|V_{t-1} = j, \cancel{\Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}}\right) \cdot p\left(V_{t-1} = j|\Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)$$

Step 3a3: from $p\left(V_t = k|\Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)$ to $p\left(V_t = k|\widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)$

$$p\left(V_t = k|\widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right) = \frac{f\left(\Delta c_t|V_t = k, \Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}\right) \cdot p\left(V_t = k|\Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)}{f\left(\Delta c_t|\Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)}$$

$$f\left(\Delta c_t|\Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}\right) = \sum_k f\left(\Delta c_t|V_t = k, \Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}\right) \cdot p\left(V_t = k|\Delta\widetilde{c_{1..t-1}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)$$

Step 3a4: Store $\boldsymbol{p\left(V_t = k|\Delta\widetilde{c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)}$ as an intermediate output of Hamilton's filter.

[15] The proof is similar to the one given in Annex F.

Restart at step 3a2 as long as $t \leq T-1$.

<u>Step 3b</u>: from $p\left(V_t|\widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)$ to $p\left(V_t|V_{t+1}, \widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)$

$$p\left(V_t|V_{t+1}, \widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right) = \frac{p\left(V_{t+1}|V_t, \widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right) \cdot p\left(V_t|\widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)}{p\left(V_{t+1}|\widetilde{\Delta c_t}, \tilde{\theta}\right)}$$

$$= \frac{p(V_{t+1}|V_t) \cdot \boldsymbol{p\left(V_t|\widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)}}{\sum_j p\left(V_{t+1}|V_t = j, \cancel{\widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}}\right) \cdot p\left(V_t = j|\widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)}$$

$\boldsymbol{p\left(V_t|\widetilde{\Delta c_{1..t}}, \widetilde{S_{1..T}}, \tilde{\theta}\right)}$ is an intermediate output of Hamilton's filter (Step 3a). A random number drawn from a uniform distribution over [0,1] is used to generate V_t based on the above formula.

<u>Step 4</u>: For j=1..4, **draw $\boldsymbol{\gamma_j^m}$** conditional on $\widetilde{\Delta y_{j,1..T}^{m,*}}$, $\widetilde{\Delta c_{j,1..T}^*}$ and σ_j^m, based on representation 2 and the 4 following independent equations:

$$\begin{cases} \Delta y_{1t}^{m,*} = \gamma_{10}^m \Delta c_{1t}^* + \sigma_1^m \cdot \varepsilon_{1t} \\ \Delta y_{2t}^{m,*} = \gamma_{20}^m \Delta c_{2t}^* + \sigma_2^m \cdot \varepsilon_{2t} \\ \Delta y_{3t}^{m,*} = \gamma_{30}^m \Delta c_{3t}^* + \sigma_3^m \cdot \varepsilon_{3t} \\ \Delta y_{4t}^{m,*} = \gamma_{40}^m \Delta c_{4t}^* + \gamma_{41}^m \Delta c_{4,t-1}^* + \gamma_{42}^m \Delta c_{4,t-2}^* + \gamma_{43}^m \Delta c_{4,t-3}^* + \sigma_4^m \cdot \varepsilon_{4t} \end{cases}$$

<u>Prior distributions</u>: $\gamma_{j0}^m \sim N\left(\underline{\gamma_j^m}, \underline{A_j^m}\right)$ for j = 1...3 and $\gamma_4^m \equiv \begin{pmatrix} \gamma_{40}^m \\ \gamma_{41}^m \\ \gamma_{42}^m \\ \gamma_{43}^m \end{pmatrix} \sim N\left(\underline{\gamma_4^m}, \underline{A_4^m}\right)$

Following Kim and Nelson (1998), we use $\underline{\gamma_j^m} = 0$ and $\underline{A_j^m} = 1$ for j = 1...3, $\underline{\gamma_4^m} = 0_{4x1}$ and $\underline{A_4^m} = Id_4$.

Notation: $\widetilde{\Delta C_{j,1..T}^*}$ is the (TxN$_j$) vector of right-hand-side variables in the above equations.

<u>Posterior distributions</u>:

$$N\left(\left[\left(\underline{A_j^m}\right)^{-1} + \left(\sigma_j^m\right)^{-2} \widetilde{\Delta C_{j,1..T}^*}\,' \widetilde{\Delta C_{j,1..T}^*}\right]^{-1} \left[\left(\underline{A_j^m}\right)^{-1} \underline{\gamma_j^m} + \left(\sigma_j^m\right)^{-2} \widetilde{\Delta C_{j,1..T}^*}\,' \widetilde{\Delta y_{j,1..T}^*}\right],\right.$$
$$\left.\left[\left(\underline{A_j^m}\right)^{-1} + \left(\sigma_j^m\right)^{-2} \widetilde{\Delta C_{j,1..T}^*}\,' \widetilde{\Delta C_{j,1..T}^*}\right]^{-1}\right)$$

<u>Step 5a</u>: **Draw $\boldsymbol{\psi_1^q}$** conditional on $\widetilde{u_{1,1..T}^q}$ and σ_1^q, based on the following equation:

$$u_{1,t}^q = \psi_{11}^q u_{1,t-1}^q + \sigma_1^q \cdot \varepsilon_{1t}^q$$

<u>Prior distribution</u>: $\psi_1^q \sim N\left(\underline{\psi_1^q}, \underline{B_1^q}\right)$. Following Kim and Nelson (1998), we use $\underline{\psi_1^q} = 0$ and $\underline{B_1^q} = 1$.

Notations: $\widetilde{u_{1,1..T}^q} \equiv \begin{pmatrix} u_{1,1}^q \\ \vdots \\ u_{1,T}^q \end{pmatrix}$

<u>Posterior distribution</u>:

$$N\left(\left[\left(\underline{B_1^q}\right)^{-1} + \left(\sigma_1^q\right)^{-2} \widetilde{u_{1,1..T-1}^q}\,' \widetilde{u_{1,1..T-1}^q}\right]^{-1} \left[\left(\underline{B_1^q}\right)^{-1} \underline{\psi_1^q} + \left(\sigma_1^q\right)^{-2} \widetilde{u_{1,1..T-1}^q}\,' \widetilde{u_{1,2..T}^q}\right],\right.$$
$$\left.\left[\left(\underline{B_1^q}\right)^{-1} + \left(\sigma_1^q\right)^{-2} \widetilde{u_{1,1..T-1}^q}\,' \widetilde{u_{1,1..T-1}^q}\right]^{-1}\right)$$

<u>Step 5b</u>: For j = 1...4, **draw $\boldsymbol{\psi_j^m}$** conditional on $\widetilde{\Delta y_{j,1..T}^m}$, $\widetilde{\Delta c_{1..T}}$ and σ_j^m, based on representation 2 and the 4 following independent equations:

$$\underbrace{\Delta y_{j,t}^m - \gamma_j^m(L)\Delta c_t}_{\equiv X_{j,t}^m} = \underbrace{\psi_{j1}^m\left(\Delta y_{j,t-1}^m - \gamma_j^m(L)\Delta c_{t-1}\right)}_{\equiv X_{j,t-1}^m \cdot \psi_j^m} + \sigma_j^m \cdot \varepsilon_{jt}$$

<u>Prior distributions</u>: $\psi_j^m \sim N\left(\underline{\psi_j^m}, \underline{B_j^m}\right)$. Following Kim and Nelson (1998), we use $\underline{\psi_j^m} = 0$ and $\underline{B_j^m} = 1$.

Notations:[16] $X_{j,t}^m \equiv \Delta y_{j,t}^m - \gamma_j^m(L)\Delta c_t$; $\widetilde{X_{j,1..T}^m} \equiv \begin{pmatrix} \Delta y_{j,1}^m - \gamma_j^m(L)\Delta c_1 \\ \vdots \\ \Delta y_{j,T}^m - \gamma_j^m(L)\Delta c_T \end{pmatrix}$

<u>Posterior distributions</u>:

$$N\left(\left[\left(\underline{B_j^m}\right)^{-1} + \left(\sigma_j^m\right)^{-2} \widetilde{X_{j,1..T-1}^m}\,' \widetilde{X_{j,1..T-1}^m}\right]^{-1} \left[\left(\underline{B_j^m}\right)^{-1} \underline{\psi_j^m} + \left(\sigma_j^m\right)^{-2} \widetilde{X_{j,1..T-1}^m}\,' \widetilde{X_{j,2..T}^m}\right],\right.$$
$$\left.\left[\left(\underline{B_j^m}\right)^{-1} + \left(\sigma_j^m\right)^{-2} \widetilde{X_{j,1..T-1}^m}\,' \widetilde{X_{j,1..T-1}^m}\right]^{-1}\right)$$

[16] Note that α_3 drawn at Step 1 includes values for Δc_3, Δc_2, Δc_1, and Δc_0, hence the possibility to compile $\gamma_j^m(L)\Delta c_3$ even for j = 4.

<u>Step 6a</u>: **Draw $\boldsymbol{\sigma_1^q}$** conditional on $\widetilde{u_{1,1..T}^q}$ and ψ_1^q, based on the following equation:

$$u_{1,t}^q = \psi_1^q u_{1,t-1}^q + \sigma_1^q \cdot \varepsilon_{1t}^q$$

<u>Prior distribution</u>: $\left(\sigma_1^q\right)^2 \sim IG\left(\nu_1^q, f_1^q\right)$. For the shape parameter, we use $\nu_1^q = 2$, the minimum required for the prior to be proper and to have a well-defined expectation, but at the same time ensuring that it remains uninformative. We choose the scale parameter f_1^q so that the expected prior variance of the idiosyncratic shock corresponds to half of the variance of GDP growth. Note that with $\nu_1^q = 2$, this prior has an expectation equal to f_1^q.

<u>Posterior distribution</u>:

$$\left(\sigma_1^q\right)^2 \sim IG\left(\nu_1^q + \frac{T-1}{2}, f_1^q + \frac{1}{2}\left(\widetilde{u_{1,2..T}^q} - \widetilde{u_{1,1..T-1}^q}\psi_1^q\right)'\left(\widetilde{u_{1,2..T}^q} - \widetilde{X_{1,1..T-1}^q}\psi_1^q\right)\right)$$

<u>Step 6b</u>: For j = 1...4, **draw $\boldsymbol{\sigma_j^m}$** conditional on $\widetilde{\Delta y_{j,1..T}^m}$, $\widetilde{\Delta c_{1..T}}$ and ψ_j^m, based on representation 2 the 4 following independent equations:

$$\underbrace{\Delta y_{j,t}^m - \gamma_j^m(L)\Delta c_t}_{\equiv X_{j,t}^m} = \underbrace{\psi_j^m\left(\Delta y_{j,t-1}^m - \gamma_j^m(L)\Delta c_{t-1}\right)}_{\equiv X_{j,t-1}^m \cdot \psi_j^m} + \sigma_j^m \cdot \varepsilon_{jt}$$

<u>Prior distributions</u>: $\left(\sigma_j^m\right)^2 \sim IG\left(\nu_j^m, f_j^m\right)$. For the shape parameters, we use $\nu_j^m = 2$, the minimum required for the priors to be proper and to have a well-defined expectation, but at the same time ensuring that they remains uninformative. We choose the scale parameters f_j^m so that the expected prior variance of the idiosyncratic shock corresponds to the variance of the corresponding observed series. Note that with $\nu_j^m = 2$, these priors have an expectation equal to f_j^m.

<u>Posterior distributions</u>:

$$\left(\sigma_j^m\right)^2 \sim IG\left(\nu_j^m + \frac{T-1}{2}, f_j^m + \frac{1}{2}\left(\widetilde{X_{j,2..T}^m} - \widetilde{X_{j,1..T-1}^m}\psi_j^m\right)'\left(\widetilde{X_{j,2..T}^m} - \widetilde{X_{j,1..T-1}^m}\psi_j^m\right)\right)$$

<u>Step 7</u>: **Draw $\boldsymbol{\phi_1}$** conditional on $\widetilde{\Delta c_{1..T}}$, $\widetilde{S_{1..T}}$, $\widetilde{V_{1..T}}$, and $\left(\mu_{S_1,V_1}, \dots, \mu_{S_T,V_T}\right)$, based on the following (state) equation:

$$\underbrace{\frac{\Delta c_t - \mu_{S_t,V_t}}{\sqrt{1 + h \cdot V_t} \cdot \sigma_c}}_{\equiv Z_t} = \phi_1 \underbrace{\frac{\Delta c_{t-1}}{\sqrt{1 + h \cdot V_t} \cdot \sigma_c}}_{\equiv Z_{t-1}} + v_t$$

<u>Prior distribution</u>: $\phi_1 \sim N\left(\underline{\phi_1}, \underline{C}\right)$. Following Kim and Nelson (1998), we use $\underline{\phi_1} = 0$ and $\underline{C} = 1$. This prior allows for factor autocorrelation but remains diffuse given its large variance.

Posterior distribution:

$$\phi_1|\widetilde{\Delta c_T},\widetilde{S_T},\mu_0,\mu_1 \sim N\left(\left(\underline{C}^{-1}+\widetilde{Z_{1..T-1}}'\widetilde{Z_{1..T-1}}\right)^{-1}\left(\underline{C}^{-1}\underline{\phi_1}+\widetilde{Z_{1..T-1}}'\widetilde{Z_{2..T}}\right),\left(\underline{C}^{-1}+\widetilde{Z_{1..T-1}}'\widetilde{Z_{1..T-1}}\right)^{-1}\right)$$

Step 8: **Draw $\boldsymbol{\mu_{00}}, \boldsymbol{\mu_{01}}, \boldsymbol{\mu_{10}}$ and $\boldsymbol{\mu_{11}}$** conditional on $\widetilde{\Delta c_{1..T}}$, $\widetilde{S_{1..T}}$, $\widetilde{V_{1..T}}$, ϕ_1, h and σ_c^2, based on the following (state) equation:

$$\frac{\Delta c_t-\phi_1\Delta c_{t-1}}{\sqrt{1+h\cdot V_t}\cdot\sigma_c} = \underbrace{\mu_{00}\frac{1}{\sqrt{1+h\cdot V_t}\cdot\sigma_c}+\mu_{01}\frac{V_t}{\sqrt{1+h\cdot V_t}\cdot\sigma_c}+\mu_{10}\frac{S_t}{\sqrt{1+h\cdot V_t}\cdot\sigma_c}+\mu_{11}\frac{S_t\cdot V_t}{\sqrt{1+h\cdot V_t}\cdot\sigma_c}}_{=\begin{pmatrix}\frac{1}{\sqrt{1+h\cdot V_t}\cdot\sigma_c} & \frac{V_t}{\sqrt{1+h\cdot V_t}\cdot\sigma_c} & \frac{S_t}{\sqrt{1+h\cdot V_t}\cdot\sigma_c} & \frac{S_t\cdot V_t}{\sqrt{1+h\cdot V_t}\cdot\sigma_c}\end{pmatrix}\begin{pmatrix}\mu_{00}\\ \mu_{01}\\ \mu_{10}\\ \mu_{11}\end{pmatrix}}+v_t$$

Prior distribution: $\tilde{\mu}\equiv\begin{pmatrix}\mu_{00}\\ \mu_{01}\\ \mu_{10}\\ \mu_{11}\end{pmatrix}\sim N\left(\underline{\tilde{\mu}},\underline{D}\right)$.

Following Kim and Nelson (1998), we use $\underline{\tilde{\mu}}=0_{4x1}$ and $\underline{D}=Id_4$.

Notations: $L_t\equiv\frac{\Delta c_t}{\sqrt{1+h\cdot V_t}\cdot\sigma_c}$; $R_t\equiv\begin{pmatrix}\frac{1}{\sqrt{1+h\cdot V_t}\cdot\sigma_c} & \frac{V_t}{\sqrt{1+h\cdot V_t}\cdot\sigma_c} & \frac{S_t}{\sqrt{1+h\cdot V_t}\cdot\sigma_c} & \frac{S_t\cdot V_t}{\sqrt{1+h\cdot V_t}\cdot\sigma_c}\end{pmatrix}$

Posterior distribution:

$$\tilde{\mu}|\widetilde{\Delta c_{1..T}},\widetilde{S_{1..T}},\widetilde{V_{1..T}},\phi_1,h,\sigma_c^2 \sim$$

$$N\left(\left(\underline{D}^{-1}+\widetilde{R_{1..T}}'\widetilde{R_{1..T}}\right)^{-1}\left(\underline{D}^{-1}\underline{\tilde{\mu}}+\widetilde{R_{1..T}}'\widetilde{L_{1..T}}\right),\left(\underline{D}^{-1}+\widetilde{R_{1..T}}'\widetilde{R_{1..T}}\right)^{-1}\right)_{I[\mu_{00}<0,\mu_{01}<0,\mu_{11}>0]}$$

Step 9: **Draw $\boldsymbol{h}$** conditional on $\widetilde{\Delta c_{1..T}}$, $\widetilde{S_{1..T}}$, $\widetilde{V_{1..T}}$, ϕ_1, $\tilde{\mu}$ and σ_c^2, based on the following (state) equation:

$$\frac{\Delta c_t}{\sigma_c}=\frac{\mu_{S_t,V_t}}{\sigma_c}+\phi_1\frac{\Delta c_{t-1}}{\sigma_c}+\sqrt{1+h\cdot V_t}\cdot v_t$$

Prior distribution: $1+h+h\,IG(\nu_h,f_h)$. For the shape parameter, we use $\nu_h=5$, thus ensuring that the prior is proper and has a well-defined expectation but at the same time remains uninformative. For the scale parameter, we use $f_h=10$, thus implying a prior expectation of 2.5.

Posterior distribution:

$$1+h \mid \widetilde{\Delta c_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \phi_1, \tilde{\mu}, \sigma_c \sim IG\left(\begin{array}{c} \nu_h + \frac{\sum_t 1_{\{V_t=1\}}}{2}, \\ f_h + \frac{\sum_t 1_{\{V_t=1\}} \cdot \left(\Delta c_t - \mu_{S_t,V_t} - \phi_1 \Delta c_{t-1}\right)^2 / \sigma_c^2}{2} \end{array}\right)_{I[1+h\geq 1]}$$

<u>Step 10</u>: Draw $\boldsymbol{\sigma_c^2}$ conditional on $\widetilde{\Delta c_{1..T}}$, $\widetilde{S_{1..T}}$, $\widetilde{V_{1..T}}$, ϕ_1, $\tilde{\mu}$ and h, based on the following (state) equation:

$$\frac{\Delta c_t}{\sqrt{1+h \cdot V_t}} = \frac{\mu_{S_t,V_t}}{\sqrt{1+h \cdot V_t}} + \phi_1 \frac{\Delta c_{t-1}}{\sqrt{1+h \cdot V_t}} + \sigma_c \cdot v_t$$

<u>Prior distribution</u>: $\sigma_c^2 \sim IG(\nu_{\sigma_c}, f_{\sigma_c})$. For the shape parameter, we use $\nu_{\sigma_c} = 2$, the minimum required for the prior to be proper and to have a well-defined expectation, but at the same time ensuring that it remains uninformative. We choose the scale parameter f_{σ_c} so that the expected prior variance of the shock to the common component corresponds to half of the variance of GDP growth (see Step 6a).

<u>Posterior distribution</u>:

$$\sigma_c^2 \mid \widetilde{\Delta c_{1..T}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \phi_1, \tilde{\mu}, h \sim IG\left(\nu_{\sigma_c} + \frac{T}{2}, f_{\sigma_c} + \frac{\sum_t \left(\frac{\Delta c_t - \mu_{S_t,V_t} - \phi_1 \Delta c_{t-1}}{\sqrt{1+h \cdot V_t}}\right)^2}{2}\right)$$

<u>Step 11</u>: Draw $\boldsymbol{\sigma_a^2}$ conditional on $\widetilde{\alpha_{1..T}}$ based on the following equation:

$$a_{1,t}^q = a_{1,t-1}^q + \sigma_a \cdot \eta_t^a$$

<u>Prior distribution</u>: $\sigma_a^2 \sim IG(\nu_{\sigma_a}, f_{\sigma_a})$. For the shape parameter, we use $\nu_{\sigma_a} = 2$, the minimum required for the prior to be proper and to have a well-defined expectation, but at the same time ensuring that it remains uninformative. We choose a scale parameter f_{σ_a} equal to 0.0005 so that the uncertainty on annual GDP growth after 10 years, as measured by its standard deviation, is roughly 2 percentage points, which is in line with typical variations over 10 years of the US potential GDP growth rate released by the Congressional Budget Office. [17] Note that the rationale is the same as in Antolin-Diaz *et al.* (2017), but our scale parameter needs to be lower than theirs because they annualise quarterly GDP growth in their information set, whereas we do not.

[17] In the FRED economic database of the Federal Reserve of Saint Louis, the corresponding series has the code GDPPOT.

Posterior distribution:

$$\sigma_a^2 \mid \widetilde{a_{1..T}} \sim IG\left(\nu_{\sigma_a} + \frac{T}{2}, f_{\sigma_a} + \frac{\sum_t \left(a_{1,t}^q - a_{1,t-1}^q\right)^2}{2}\right)$$

Step 12: **Draw transition probabilities $\boldsymbol{PrS_{11}}$ and $\boldsymbol{PrS_{00}}$** conditional on $\widetilde{S_{1..T}}$

Prior distributions: $PrS_{11} \sim beta\left(u_{11}^S, u_{10}^S\right)$; $PrS_{00} \sim beta\left(u_{00}^S, u_{01}^S\right)$. Following Kim and Nelson (1998), we rely on informative priors reflecting the number and length of recessions over the estimation sample: $u_{00}^S = 90$; $u_{01}^S = u_{10}^S = 7$; $u_{11}^S = 470$. The NBER estimates that the US economy has known seven recession episodes between January 1970 and December 2017, covering 90 months in total.

Likelihood function: $L\left(PrS_{11}, PrS_{00} | \widetilde{S_{1..T}}\right) = {PrS_{11}}^{n_{11}^S}(1 - PrS_{11})^{n_{10}^S}{PrS_{00}}^{n_{00}^S}(1 - PrS_{00})^{n_{01}^S}$ where n_{ij}^S refer to transitions from state i to state j which can be counted based on $\widetilde{S_T}$.

Posterior distributions:

$$PrS_{11} | \widetilde{S_{1..T}} \sim beta\left(u_{11}^S + n_{11}^S, u_{10}^S + n_{10}^S\right) \;\; ; \;\; PrS_{00} | \widetilde{S_{1..T}} \sim beta\left(u_{00}^S + n_{00}^S, u_{01}^S + n_{01}^S\right)$$

Step 13: **Draw transition probabilities $\boldsymbol{PrV_{11}}$ and $\boldsymbol{PrV_{00}}$** conditional on $\widetilde{V_{1..T}}$

Prior distributions: $PrV_{11} \sim beta(u_{11}^V, u_{10}^V)$; $PrV_{00} \sim beta(u_{00}^V, u_{01}^V)$. We rely on informative priors to guide the identification of volatility states: $u_{00}^S = 285$; $u_{01}^S = u_{10}^S = 1$; $u_{11}^S = 285$. These priors correspond to a situation where the economy would spend roughly half of the time in a high-volatility state and half of the time in a low-volatility state, with a limited number of transitions between the two states.

Likelihood function: $L\left(PrV_{11}, PrV_{00} | \widetilde{V_{1..T}}\right) = {PrV_{11}}^{n_{11}^V}(1 - PrV_{11})^{n_{10}^V}{PrV_{00}}^{n_{00}^V}(1 - PrV_{00})^{n_{01}^V}$ where n_{ij}^V refer to transitions from state i to state j which can be counted based on $\widetilde{V_T}$.

Posterior distributions:

$$PrV_{11} | \widetilde{V_{1..T}} \sim beta(u_{11}^V + n_{11}^V, u_{10}^V + n_{10}^V) \;\; ; \;\; PrV_{00} | \widetilde{V_{1..T}} \sim beta(u_{00}^V + n_{00}^V, u_{01}^V + n_{01}^V)$$

Annex F. Smoothed inference on S_t (business cycle state: expansion or recession)

The state equation governing the dynamics of the underlying factor is given by:

$$\phi(L)\Delta c_t = \mu_{S_t,V_t} + \sqrt{1 + h \cdot V_t} \cdot \sigma_c \cdot \eta_t^c \quad ; \quad \eta_t^c \sim N(0,1)$$

<u>Result 1</u>: $\boldsymbol{p\left(\widetilde{S_{1..T}} \middle| \widetilde{\Delta y^*_{1..T}}, \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) = p\left(\widetilde{S_{1..T}} \middle| \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$

Proof: Once $\widetilde{\Delta c_{1..T}}$ is included in the information, $\widetilde{\Delta y^*_{1..T}}$ does not bring any additional information on $\widetilde{S_{1..T}}$.

$p\left(\widetilde{S_{1..T}} \middle| \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$ can then be rewritten as follows:

$$p\left(\widetilde{S_{1..T}} \middle| \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) = p\left(S_T \middle| \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) \cdot p\left(S_{T-1} \middle| S_T, \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) \cdot p\left(\widetilde{S_{1..T-2}} \middle| S_{T-1}, S_T, \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$$

<u>Result 2</u>: $\boldsymbol{p\left(S_{T-1} \middle| S_T, \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) = p\left(S_{T-1} \middle| S_T, \widetilde{\Delta c_{1..T-1}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$

Proof:

$$p\left(S_{T-1} \middle| S_T, \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$$

$$= \frac{p\left(\Delta c_T \middle| S_{T-1}, S_T, \widetilde{\Delta c_{1..T-1}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) \cdot p\left(S_{T-1} \middle| S_T, \widetilde{\Delta c_{1..T-1}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}{p\left(\Delta c_T \middle| S_T, \widetilde{\Delta c_{1..T-1}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$$

$$= p\left(S_{T-1} \middle| S_T, \widetilde{\Delta c_{1..T-1}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$$

The last equality follows from the fact that Δc_T does not depend on S_{T-1}, hence:

$$p\left(\Delta c_T \middle| S_{T-1}, S_T, \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) = p\left(\Delta c_T \middle| S_T, \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$$

<u>Result 3</u>: $\boldsymbol{p\left(\widetilde{S_{1..T-2}} \middle| S_{T-1}, S_T, \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) = p\left(\widetilde{S_{1..T-2}} \middle| S_{T-1}, S_T, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$

Proof:

$$p\left(\widetilde{S_{1..T-2}} \middle| S_{T-1}, S_T, \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$$

$$= \frac{p\left(\widetilde{S_{1..T-2}}, \Delta c_{T-1}, \Delta c_T \middle| S_{T-1}, S_T, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}{p\left(\Delta c_{T-1}, \Delta c_T \middle| S_{T-1}, S_T, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$$

$$= \frac{p\left(\Delta c_{T-1}, \Delta c_T \middle| \widetilde{S_{1..T-2}}, S_{T-1}, S_T, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) \cdot p\left(\widetilde{S_{1..T-2}} \middle| S_{T-1}, S_T, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}{p\left(\Delta c_{T-1}, \Delta c_T \middle| S_{T-1}, S_T, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_T}, \tilde{\theta}\right)}$$

$$= p\left(\widetilde{S_{1..T-2}} \middle| S_{T-1}, S_T, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$$

The last equality follows from the fact that Δc_{T-1} and Δc_T do not depend on $\widetilde{S_{1..T-2}}$, hence:

$$p\left(\Delta c_{T-1}, \Delta c_T | \widetilde{S_{1..T-2}}, S_{T-1}, S_T, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) = p\left(\Delta c_{T-1}, \Delta c_T | S_{T-1}, S_T, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$$

<u>Result 4</u>: $\boldsymbol{p\left(\widetilde{S_{1..T-2}} | S_{T-1}, S_T, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) = p\left(\widetilde{S_{1..T-2}} | S_{T-1}, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$

Proof:

$$p\left(\widetilde{S_{1..T-2}} | S_{T-1}, S_T, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$$

$$= \frac{p\left(S_T | \widetilde{S_{1..T-2}}, S_{T-1}, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) \cdot p\left(\widetilde{S_{1..T-2}} | S_{T-1}, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}{p\left(S_T | S_{T-1}, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$$

$$= p\left(\widetilde{S_{1..T-2}} | S_{T-1}, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$$

The last equality follows from the fact that

$$p\left(S_T | \widetilde{S_{1..T-2}}, S_{T-1}, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) = p\left(S_T | S_{T-1}, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$$

As a result, the smoothed inference on the state of the business cycle can be rewritten as follows:

$$\boxed{\begin{aligned} &\boldsymbol{p\left(\widetilde{S_{1..T}} | \widetilde{\Delta y^*_{1..T}}, \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)} \\ &\boldsymbol{= p\left(S_T | \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) \cdot p\left(S_{T-1} | S_T, \widetilde{\Delta c_{1..T-1}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) \cdot p\left(\widetilde{S_{1..T-2}} | S_{T-1}, \widetilde{\Delta c_{1..T-2}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)} \\ &\boldsymbol{= p\left(S_T | \widetilde{\Delta c_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) \cdot \prod_{t=1}^{T-1} p\left(S_t | S_{t+1}, \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)} \end{aligned}}$$

Note that $p\left(S_t = j | S_{t+1} = k, \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$ can be inferred from the filtered probabilities:

$$p\left(S_t = j | S_{t+1} = k, \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)$$

$$= \frac{p\left(S_{t+1} = k | S_t = j, \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) \cdot p\left(S_t = j | \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}{p\left(S_{t+1} = k | = \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$$

$$= \frac{p\left(S_{t+1} = k | S_t = j, \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) \cdot p\left(S_t = j | \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}{\sum_j p\left(S_{t+1} = k | S_t = j, \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right) \cdot p\left(S_t = j | \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$$

$$= \frac{p(S_{t+1} = k | S_t = j) \cdot p\left(S_t = j | \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}{\sum_j p(S_{t+1} = k | S_t = j) \cdot p\left(S_t = j | \widetilde{\Delta c_{1..t}}, \widetilde{V_{1..T}}, \tilde{\theta}\right)}$$

Annex G. Marginal likelihood computation of MS-DFMs

Marginal likelihoods allow to compare different DFM specifications (e.g. with or without Markov-Switching features). Nevertheless, these quantities are typically difficult to estimate for hierarchical models with large state vectors such as MS-DFMs.

Chib (1995) suggests to rely on the output of the Gibbs sampler in order to estimate the marginal likelihood of a model. Denoting the vector of model parameters by $\tilde{\theta}$, his method rests on the following formula which relates the marginal and conditional likelihoods of the model and the prior and posterior parameter densities:

$$f\left(\widetilde{\Delta y_{1..T}}\right) = \frac{f\left(\widetilde{\Delta y_{1..T}}|\tilde{\theta}\right) \cdot \varphi\left(\tilde{\theta}\right)}{\varphi\left(\tilde{\theta}|\widetilde{\Delta y_{1..T}}\right)} \qquad (\mathbf{1})$$

Since this formula holds for any admissible parameter value, it holds in particular for $\tilde{\theta}^*$, the posterior mean of the vector of parameters, which is often recommended when applying Chib's (1995) formula. Taking logarithms, this formula can be then rewritten as follows[18]:

$$\log f\left(\widetilde{\Delta y_{1..T}}\right) = \log f\left(\widetilde{\Delta y_{1..T}}|\tilde{\theta}^*\right) + \log \varphi\left(\tilde{\theta}^*\right) - \log \varphi\left(\tilde{\theta}^*|\widetilde{\Delta y_{1..T}}\right) \qquad (\mathbf{2})$$

Evaluating $f\left(\widetilde{\Delta y_{1..T}}|\tilde{\theta}^*\right)$

In the case of a MS-DFM, $\log f\left(\widetilde{\Delta y_{1..T}}|\tilde{\theta}^*\right)$ cannot be evaluated analytically, but the Kalman filter step of the Gibbs Sampling algorithm provides draws from another conditional log-likelihood, namely $\log f\left(\widetilde{\Delta y_{1..T}}|\widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}^*\right)$, where the conditioning set includes the two Markov-Switching variables $\widetilde{S_{1..T}}$ and $\widetilde{V_{1..T}}$, in addition to the vector of model parameters. These two conditional log-likelihoods can be decomposed into sums of one-step ahead prediction densities, as follows:

$$\log f\left(\widetilde{\Delta y_{1..T}}|\tilde{\theta}^*\right) = \sum_{t=1}^{T} \log f\left(\Delta y_t|\widetilde{\Delta y_{1..t-1}}, \tilde{\theta}^*\right) \qquad (\mathbf{3})$$

$$\log f\left(\widetilde{\Delta y_{1..T}}|\widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}^*\right) = \sum_{t=1}^{T} \log f\left(\Delta y_t|\widetilde{\Delta y_{1..t-1}}, \widetilde{S_{1..T}}, \widetilde{V_{1..T}}, \tilde{\theta}^*\right)$$

$$= \sum_{t=1}^{T} \log f\left(\Delta y_t|\widetilde{\Delta y_{1..t-1}}, \widetilde{S_{1..t}}, \widetilde{V_{1..t}}, \tilde{\theta}^*\right) \quad (\mathbf{4})$$

For each date, $f\left(\Delta y_t|\widetilde{\Delta y_{1..t-1}}, \tilde{\theta}^*\right)$ is related to $f\left(\Delta y_t|\widetilde{\Delta y_{1..t-1}}, \widetilde{S_{1..t}}, \widetilde{V_{1..t}}, \tilde{\theta}^*\right)$ by the following formula:

[18] Note that $\tilde{\theta}^*$ includes all constant model parameters, but no Markov-Switching variable. It is indeed well known that Chib's (1995) formula gives more accurate estimates of model marginal likelihoods if the number of variables in the information set is limited to the maximum extent.

$$f\left(\Delta y_t|\Delta\widetilde{y_{1..t-1}},\tilde{\theta}^*\right)=\int f\left(\Delta y_t|\Delta\widetilde{y_{1..t-1}},\widetilde{S_{1..t}},\widetilde{V_{1..t}},\tilde{\theta}^*\right)\cdot g\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\Delta\widetilde{y_{1..t-1}},\tilde{\theta}^*\right)\cdot d\widetilde{S_{1..t}}\cdot d\widetilde{V_{1..t}} \qquad (\mathbf{5})$$

At first sight, it might seem that this integral cannot be evaluated because $\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\Delta\widetilde{y_{1..t-1}},\tilde{\theta}^*\right)$ is not known analytically, contrary to $f\left(\Delta y_t|\Delta\widetilde{y_{1..t-1}},\widetilde{S_{1..t}},\widetilde{V_{1..t}},\tilde{\theta}^*\right)$ which is an output of the Kalman filter. Nevertheless, the integral can be numerically approximated by drawing from $g\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\Delta\widetilde{y_{1..t-1}},\tilde{\theta}^*\right)$ using a particle filter.

Particle filter

The particle filter applies the logic of sequential importance sampling.[19] Instead of drawing from $g\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\theta}^*\right)$ independently at each date, what the algorithm developed by Gerlach, Carter and Kohn (2000) would allow doing at very high computational costs if were used for as many dates as in the present setup where the sample covers close to 600 months, $g\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\theta}^*\right)$ is approximated with an importance distribution $q_{1..t}\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\psi}\right)$ taking a recursive form:

$$q_{1..t}\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\psi}\right)\equiv q_t\left(S_t,V_t|\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}},\widetilde{\Delta y_{1..t}},\tilde{\psi}\right)\cdot q_{1..t-1}\left(\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}}|\Delta\widetilde{y_{1..t-1}},\tilde{\psi}\right) \qquad (\mathbf{6})$$

Particle weights

Since particles are drawn from the importance distribution $q_{1..t}\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\psi}\right)$ rather than the target distribution $g\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\theta}^*\right)$, they need to be appropriately weighted for the evaluation of $f\left(\Delta y_t|\Delta\widetilde{y_{1..t-1}},\tilde{\theta}^*\right)$. Equation (5) can be rewritten as follows:

$$f\left(\Delta y_t|\Delta\widetilde{y_{1..t-1}},\tilde{\theta}^*\right)=\int f\left(\Delta y_t|\Delta\widetilde{y_{1..t-1}},\widetilde{S_{1..t}},\widetilde{V_{1..t}},\tilde{\theta}^*\right)\cdot g\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\Delta\widetilde{y_{1..t-1}},\tilde{\theta}^*\right)\cdot d\widetilde{S_{1..t}}\cdot d\widetilde{V_{1..t}}$$

$$=\int f\left(\Delta y_t|\Delta\widetilde{y_{1..t-1}},\widetilde{S_{1..t}},\widetilde{V_{1..t}},\tilde{\theta}^*\right)\cdot p\left(S_t,V_t|S_{t-1},V_{t-1},\tilde{\theta}^*\right)\cdot\underbrace{\frac{g\left(\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}}|\Delta\widetilde{y_{1..t-1}},\tilde{\theta}^*\right)}{q_{1..t-1}\left(\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}}|\Delta\widetilde{y_{1..t-1}},\tilde{\psi}\right)}}_{\equiv W_{t-1}\left(\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}}\right)}$$

$$\cdot\, q_{1..t-1}\left(\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}}|\Delta\widetilde{y_{1..t-1}},\tilde{\psi}\right)\cdot d\widetilde{S_{1..t}}\cdot d\widetilde{V_{1..t}} \qquad (\mathbf{7})$$

Particle weights at date t are defined as $W_t\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)\equiv\frac{g\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\theta}^*\right)}{q_{1..t}\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\psi}\right)}$. It is convenient to rewrite these weights as a ratio of un-normalised weights $w_t\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)$ to a normalisation factor Z_t, as shown in equation (8).

[19] For an introduction to particle filters and sequential importance sampling, see Doucet and Johansen (2011), and in particular their Section 4.6.1 on conditionally linear Gaussian models.

$$W_t\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)=\underbrace{\frac{h\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}},\widetilde{\Delta y_{1..t}}|\tilde{\theta}^*\right)}{q_{1..t}\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\psi}\right)}}_{\equiv w_t\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)}\cdot\underbrace{\frac{1}{\int h\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}},\widetilde{\Delta y_{1..t}}|\tilde{\theta}^*\right)\cdot d\widetilde{S_{1..t}}\cdot d\widetilde{V_{1..t}}}}_{\equiv\frac{1}{Z_t}} \qquad (\mathbf{8})$$

The normalisation factor Z_t can itself be computed using the importance distribution $q_{1..t}\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\theta}^*\right)$:

$$\begin{aligned}Z_t &\equiv \int h\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}},\widetilde{\Delta y_{1..t}}|\tilde{\theta}^*\right)\cdot d\widetilde{S_{1..t}}\cdot d\widetilde{V_{1..t}}\\ &= \int \underbrace{\frac{h\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}},\widetilde{\Delta y_{1..t}}|\tilde{\theta}^*\right)}{q_{1..t}\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\psi}\right)}}_{\equiv w_t\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)}\cdot q_{1..t}\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\psi}\right)\cdot d\widetilde{S_{1..t}}\cdot d\widetilde{V_{1..t}} \qquad (\mathbf{9})\end{aligned}$$

Equation (9) shows that, in practice, Z_t can be computed as the sum of unnormalised weights for a sample of particles drawn according to the importance distribution $q_{1..t}\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}|\widetilde{\Delta y_{1..t}},\tilde{\theta}^*\right)$. In the following, we will thus focus on unnormalised weights.

It turns out that a convenient recursion applies for the computation of un-normalised particle weights, as shown by equation (10).

$$\begin{aligned}&w_t\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)\\ &=\underbrace{\frac{h\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}},\widetilde{\Delta y_{1..t}}|\tilde{\theta}^*\right)}{h\left(\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}},\widetilde{\Delta y_{1..t-1}}|\tilde{\theta}^*\right)\cdot q_t\left(S_t,V_t|\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}},\widetilde{\Delta y_{1..t}},\tilde{\psi}\right)}}_{\equiv\alpha_t\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)}\ \underbrace{\frac{h\left(\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}},\widetilde{\Delta y_{1..t-1}}|\tilde{\theta}^*\right)}{q_{1..t-1}\left(\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}}|\widetilde{\Delta y_{1..t-1}},\tilde{\psi}\right)}}_{\equiv w_{t-1}\left(\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}}\right)}\end{aligned} \qquad (\mathbf{10})$$

The transition factor $\alpha_t\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)$ between un-normalised weights at dates (t-1) and t can be rewritten as follows:

$$\alpha_t\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)=\frac{p\left(S_t,V_t,\Delta y_t|\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}},\widetilde{\Delta y_{1..t-1}},\tilde{\theta}^*\right)}{q_t\left(S_t,V_t|\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}},\widetilde{\Delta y_{1..t}},\tilde{\psi}\right)}$$

$$=\frac{f\left(\Delta y_t|\widetilde{\Delta y_{1..t-1}},\widetilde{S_{1..t}},\widetilde{V_{1..t}},\tilde{\theta}^*\right)\cdot p\left(S_t,V_t|\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}},\widetilde{\Delta y_{1..t-1}},\tilde{\theta}^*\right)}{q_t\left(S_t,V_t|\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}},\widetilde{\Delta y_{1..t}},\tilde{\psi}\right)}$$

$$\Rightarrow \alpha_t\left(\widetilde{S_{1..t}},\widetilde{V_{1..t}}\right)=\frac{f\left(\Delta y_t|\widetilde{\Delta y_{1..t-1}},\widetilde{S_{1..t}},\widetilde{V_{1..t}},\tilde{\theta}^*\right)\cdot p\left(S_t,V_t|S_{t-1},V_{t-1},\tilde{\theta}^*\right)}{q_t\left(S_t,V_t|\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}},\widetilde{\Delta y_{1..t}},\tilde{\psi}\right)} \qquad (\mathbf{11})$$

The last equality follows from the Markov structure of the model, implying $p\left(S_t,V_t|\widetilde{S_{1..t-1}},\widetilde{V_{1..t-1}},\widetilde{\Delta y_{1..t-1}},\tilde{\theta}^*\right)=p\left(S_t,V_t|S_{t-1},V_{t-1},\tilde{\theta}^*\right)$.

Resampling

In the present setting where the time horizon is long (close to 600 months), it is very likely that some particles get a disproportionate weight as compared to others, which leads to an increasing variance of all particle-based estimators. In order to avoid this problem, we apply multinomial resampling each time the effective sample size (ESS) falls below half of the sample size. We thus follow Doucet and Johansen (2011) who report that multinomial resampling outperforms other resampling schemes in most scenarios and that it is better not to resample at each step. In practice, the effective sample size at date t for a sample of N particles indexed by *i* is computed as follows:

$$ESS_t = \left(\sum_{i=1}^{N} W_{it}^2\right)^{-1} \qquad (\mathbf{12})$$

It takes values between 1 (when a single particle gets all the weight) and N (when particles are equally weighted).

Functional form of the importance sampling distribution

We follow Doucet and Johansen (2011, Section 4.6.1) for the choice of the importance sampling distribution that minimises the variance of particle weights at each date:

$$q_t\left(S_t, V_t | \widetilde{S_{1..t-1}}, \widetilde{V_{1..t-1}}, \widetilde{\Delta y_{1..t}}, \tilde{\psi}\right) \equiv p\left(S_t, V_t | \widetilde{S_{1..t-1}}, \widetilde{V_{1..t-1}}, \widetilde{\Delta y_{1..t}}, \tilde{\theta}^*\right) \qquad (\mathbf{13})$$

This probability can be evaluated for the four possible values that (S_t, V_t) can take, using Bayes formula and the estimation of $f\left(\Delta y_t | \widetilde{\Delta y_{1..t-1}}, \widetilde{S_{1..t}}, \widetilde{V_{1..t}}, \tilde{\theta}^*\right)$ returned by the Kalman filter:

$$p\left(S_t, V_t | \widetilde{S_{1..t-1}}, \widetilde{V_{1..t-1}}, \widetilde{\Delta y_{1..t}}, \tilde{\theta}^*\right) = \frac{f\left(\Delta y_t | \widetilde{\Delta y_{1..t-1}}, \widetilde{S_{1..t}}, \widetilde{V_{1..t}}, \tilde{\theta}^*\right) \cdot p\left(S_t, V_t | \widetilde{S_{1..t-1}}, \widetilde{V_{1..t-1}}, \widetilde{\Delta y_{1..t-1}}, \tilde{\theta}^*\right)}{f\left(\Delta y_t | \widetilde{\Delta y_{1..t-1}}, \widetilde{S_{1..t-1}}, \widetilde{V_{1..t-1}}, \tilde{\theta}^*\right)}$$

$$= \frac{f\left(\Delta y_t | \widetilde{\Delta y_{1..t-1}}, \widetilde{S_{1..t}}, \widetilde{V_{1..t}}, \tilde{\theta}^*\right) \cdot p\left(S_t, V_t | S_{t-1}, V_{t-1}, \tilde{\theta}^*\right)}{f\left(\Delta y_t | \widetilde{\Delta y_{1..t-1}}, \widetilde{S_{1..t-1}}, \widetilde{V_{1..t-1}}, \tilde{\theta}^*\right)} \qquad (\mathbf{14})$$

This choice of functional form for the importance sampling distribution q_t also has implications for $\alpha_t\left(\widetilde{S_{1..t}}, \widetilde{V_{1..t}}\right)$, which can be rewritten as follows:

$$\alpha_t\left(\widetilde{S_{1..t}}, \widetilde{V_{1..t}}\right) = f\left(\Delta y_t | \widetilde{\Delta y_{1..t-1}}, \widetilde{S_{1..t-1}}, \widetilde{V_{1..t-1}}, \tilde{\theta}^*\right)$$

$$= \sum_{i,j} f\left(\Delta y_t | \widetilde{\Delta y_{1..t-1}}, \widetilde{S_{1..t-1}}, \widetilde{V_{1..t-1}}, S_t = i, V_t = j, \tilde{\theta}^*\right) \cdot p\left(S_t = i, V_t = j | S_{t-1}, V_{t-1}, \tilde{\theta}^*\right) \qquad (\mathbf{15})$$

Equation (15) shows that $\alpha_t\left(\widetilde{S_{1..t}}, \widetilde{V_{1..t}}\right)$ can be evaluated using $f\left(\Delta y_t | \widetilde{\Delta y_{1..t-1}}, \widetilde{S_{1..t}}, \widetilde{V_{1..t}}, \tilde{\theta}^*\right)$, which is an output of the Kalman filter, and the prior transition probabilities of the Markov-Switching variables.

Evaluating $\varphi(\tilde{\theta}^*)$ and $\varphi(\tilde{\theta}^*|\widetilde{\Delta y_{1..T}})$

Here we follow the strategy advocated by Bauwens and Rombouts (2012). The key to evaluate $\varphi(\tilde{\theta}^*)$ and $\varphi(\tilde{\theta}^*|\widetilde{\Delta y_{1..T}})$ is to split the vector $\tilde{\theta}^*$ into the same subvectors $(\widetilde{\theta_1}, \dots, \widetilde{\theta_K})$ as in the Gibbs sampling algorithm. In this case, the prior density $\varphi(\tilde{\theta}^*)$ can be evaluated using K analytically known distributions, as follows:

$$\varphi(\tilde{\theta}^*) = \prod_{i=1}^{K} \varphi(\widetilde{\theta_i^*}) \qquad (\mathbf{16})$$

The posterior density $\varphi(\tilde{\theta}^*|\widetilde{\Delta y_{1..T}})$ can also be factorised into K terms corresponding to the different Gibbs sampling steps:

$$\varphi(\tilde{\theta}^*|\widetilde{\Delta y_{1..T}}) = \varphi(\widetilde{\theta_1^*}|\widetilde{\Delta y_{1..T}}) \cdot \varphi(\widetilde{\theta_2^*}|\widetilde{\theta_1^*}, \widetilde{\Delta y_{1..T}}) \cdot \dots \cdot \varphi(\widetilde{\theta_K^*}|\widetilde{\theta_1^*}, \dots, \widetilde{\theta_{K-1}^*}, \widetilde{\Delta y_{1..T}}) \qquad (\mathbf{17})$$

Each of these terms is finally evaluated using G draws from an auxiliary Gibbs sampler. For example, $\varphi(\widetilde{\theta_2^*}|\widetilde{\theta_1^*}, \widetilde{\Delta y_{1..T}})$ is evaluated as:

$$\frac{1}{G} \cdot \sum_{i=1}^{G} \varphi(\widetilde{\theta_2^*}|\widetilde{\theta_1^*}, \widetilde{\theta_3^l}, \dots, \widetilde{\theta_K^l}, \widetilde{S_{1..T}^l}, \widetilde{V_{1..T}^l}, \widetilde{\Delta y_{1..T}})$$

where $\widetilde{\theta_1^*}$ is the posterior mean of the first subvector of parameters, and $(\widetilde{\theta_3^l}, \dots, \widetilde{\theta_K^l}, \widetilde{S_{1..T}^l}, \widetilde{V_{1..T}^l})$ are draws from an auxiliary Gibbs sampler which is similar to the one used in the main text, except that θ_1 is set equal to $\widetilde{\theta_1^*}$. The same reasoning applies for the other terms in equation (17).