Business Cycle Dynamics after the Great Recession: An Extended Markov-Switching Dynamic Factor Model

Catherine DOZ, Laurent FERRARA and Pierre-Alain PIONNIER

31 May 2019

Preliminary Draft

Abstract:
The Great Recession and the subsequent period of subdued GDP growth in most advanced countries have highlighted that macroeconomic forecasters need to account for sudden and deep recessions, periods of higher macroeconomic volatility and low-frequency 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 main features: switches in volatility and time-variation in long-term 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. This is an important result for the detection of future recessions since, according to our model, the U.S. economy has been facing a low macroeconomic volatility since the onset of Great Moderation period, only temporarily interrupted by the Great Recession. Second, we point out that our extended model is able to capture a continuous decline in the U.S. GDP growth trend that started a few years before the Great Recession. Those two new extensions of the standard MS-DFM framework are supported by information criteria, marginal likelihood comparisons and improved real-time GDP forecasting performances.

Keywords: Markov-Switching Dynamic Factor Models, Great Moderation, Great Recession, Turning-Point Detection, Macroeconomic Forecasting

JEL Codes: C22, C51, E32, E37

1 Catherine Doz, PSE and University Paris 1 Panthéon-Sorbonne, email: catherine.doz@univ-paris1.fr
Laurent Ferrara, Banque de France, International Macro Division, email: laurent.ferrara@banque-france.fr
Pierre-Alain Pionnier, OECD, Statistics and Data Directorate, email: pierre-alain.pionnier@oecd.org

We would like to thank Guillaume Chevillon, Arnaud Dufays, Luca Sala and participants at the 2018 OxMetrics User Conference, the 2018 French Econometrics Conference, the 2019 SNDE Symposium and the Namur University and Bocconi University seminars for comments and suggestions.
1. Introduction

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 overestimate 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 its long-term average (see for example Ferrara and Marsillli, 2017, as regards IMF forecasts). In addition, changes in macroeconomic volatility over the last decade further complicated the task of forecasters. Indeed, 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, or 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 an expansion or a recession phase, are increasingly used for short-term forecasting and turning point detection (Chauvet and Potter 2012, 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 in 1946. Chauvet and Piger (2008) and Hamilton (2011) later confirmed the very good empirical performance of small-scale MS-DFMs to track U.S. recessions. They showed that MS-DFMs using the four monthly series advocated by the NBER (namely 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 U.S. recessions in a fully transparent way. In addition, they showed that such models were able to identify business cycle phases faster than the NBER, especially around troughs. These small-scale MS-DFMs also compare favourably to the univariate Markov-Switching model of Hamilton (1989), which only extracts information from the U.S. 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 in order to date business cycles.

Against this background, our contribution to the literature is to put forward a new extended MS-DFM which is able to replicate the two recent empirical stylized facts observed after the Great Recession. The first stylised fact 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 U.S. after the Great Recession. To account for the
second stylized fact, namely changing macroeconomic volatility, instead of assuming a stochastic volatility for the error term of the DFM, in the spirit of the papers by Marcellino et al. (2016) and Antolin-Diaz et al. (2017), we allow for switches in volatility in the state equation of the model, according to a Markov chain that differs from the one on the conditional mean.

Based on U.S. data, we show that the introduction of Markov-Switching volatility in the standard MS-DFM framework is supported by statistical criteria (Deviance Information Criterion and marginal likelihoods) and improves the detection of turning points after the mid-1980s. The extended MS-DFM also accurately detects the Great Recession with a lead of about 8 months compared with 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 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 U.S. data, we also get two interesting results. First, we note a decline in the U.S. long-run GDP growth rate of about 1 p.p. per year as compared to the early 2000s, half of it occurring before the Great Recession. Such decline in long-run U.S. GDP growth has been empirically documented by Antolin-Diaz et al. (2017) and rationalized by Fernald et al. (2017) who highlight the role of total factor productivity and labor 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 is characterized 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 U.S. recessions. This 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 buoyant literature on the topic. 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 U.S. data from 1970 to 2017. In particular, we show that accounting for switches in volatility improves the detection of turning points after the mid-1980s and that the Great Recession did not put an end to the Great Moderation. We also provide evidence of a decline in trend GDP growth that started a few years before the Great Recession. We finally show that our extensions of the standard MS-DFM framework are supported by information criteria and marginal likelihood comparisons. In Section 6, we evaluate the real-time properties of our model. We show that this new model leads to an improved short-term forecasting performance of U.S. quarterly GDP growth in real-time, both during and after the Great Recession. We finally notice for further research that an even earlier detection of turning points would allow further improvements in GDP forecasting performance.
2. Related literature

Markov-Switching Dynamic Factor Models (MS-DFMs) have long tradition in business cycle analysis initiated by the pioneer work of Diebold and Rudebusch (1996), thanks to their ability to simultaneously account for comovement in macroeconomic time series and changes in growth regimes. Against this background many extensions have been put forward in the literature. In this paper, we put forward an innovative extended MS-DFM that also allows for two recent stylized facts: 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 U.S. 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 by a gradual decline, reflecting thus a reduction in the size of shocks hitting the economy more than a change in the propagation mechanism of these shocks.

Our model is also 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).

We consider a larger information set than Eo and Kim (2016), 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 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 U.S. quarterly GDP. Based on this specification, they conclude that the Great Recession was U-shaped and that the U.S. economy fully recovered from it in 2014.

All three papers identify a decrease in U.S. trend GDP growth which started before the Great Recession, but its precise timing and magnitude differ. Antolin-Diaz et al. (2017) conclude that the U.S. 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 U.S. 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 U.S. 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), U.S. 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 below summarises how the present paper relates to the existing literature.

### Table 1: Comparison with the existing literature

<table>
<thead>
<tr>
<th></th>
<th>Multivariate framework (factor model)</th>
<th>Markov-Switching Intercept</th>
<th>Markov-Switching volatility of shocks</th>
<th>Stochastic volatility</th>
<th>Time-varying long-term GDP growth rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antolin-Diaz et al.</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(2017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai and Wang (2011)</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diebold and Rudebusch (1996)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eo and Kim (2016)</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Giordani et al. (2007)(^3)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim and Nelson (1998)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marcellino et al. (2016)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This paper</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

\(^2\) 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 U.S. quarterly GDP as observable variable, whereas we consider the four monthly series explicitly mentioned by the business cycle Dating Committee of the NBER, in addition to quarterly GDP.

We consider a state-space model where, for the time being, all observable variables are available at monthly frequency. This model is given by the two following equations:

\[
\begin{align*}
\text{Measurement equations (for } i = 1, \ldots, n): & \quad \Delta y_{it} = a_{i,t} + \gamma_i(L) \Delta c_t + u_{it} \\
\text{State equations:} & \quad \phi(L) \Delta c_t = \mu_{c,t} + \sqrt{1 + h \cdot V_t} \cdot \sigma_c \cdot \eta_c^i \sim IID \ N(0,1) \\
& \quad \psi_i(L) u_{it} = \sigma_i \cdot \xi_{it} \sim IID \ N(0,1)
\end{align*}
\]

In the measurement equations, the \( \Delta c_t \) 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 U.S. 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 \( \Delta c_t \) 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).

\footnote{See also Camacho et al. (2014) for an application to the euro area}
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:

\[
\mu_{S_t, V_t} \equiv \mu_{00} + \mu_{01} \cdot V_t + \mu_{10} \cdot S_t + \mu_{11} \cdot S_t \cdot V_t
\]

where

\[
\mu_{00} : \text{Intercept in the low-volatility recession regime}
\]
\[
\mu_{00} + \mu_{01} : \text{Intercept in the high-volatility recession regime}
\]
\[
\mu_{00} + \mu_{10} : \text{Intercept in the low-volatility expansion regime}
\]
\[
\mu_{00} + \mu_{01} + \mu_{10} + \mu_{11} : \text{Intercept in the high-volatility expansion regime}
\]

At this stage, it is useful to clarify how the model can accommodate a mix of monthly and quarterly observable variables. The aim is to introduce quarterly GDP in addition to monthly macroeconomic variables in the measurement equations. Quarterly GDP $\Delta y_{1t}^q$ is observed once every three months whereas monthly variables $\Delta y_{jt}^m$ 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 measurement equations can thus be rewritten as follows:

\[
\begin{align*}
\Delta y_{1t}^q &= \left(\frac{1}{3}a_{1t}^q + \frac{2}{3}a_{1,t-1}^q + a_{1,t-2}^q + \frac{2}{3}a_{1,t-3}^q + \frac{1}{3}a_{1,t-4}^q\right) + y_1^q(L)\left(\frac{1}{3}\Delta c_t + \frac{2}{3}\Delta c_{t-1} + \Delta c_{t-2} + \frac{2}{3}\Delta c_{t-3} + \frac{1}{3}\Delta c_{t-4}\right) \\
& \quad + \left(\frac{1}{3}u_{1,t}^q + \frac{2}{3}u_{1,t-1}^q + u_{1,t-2}^q + \frac{2}{3}u_{1,t-3}^q + \frac{1}{3}u_{1,t-4}^q\right) \\
\Delta y_{jt}^m &= a_{jt}^m + y_j^m(L)\Delta c_t + u_{jt}^m
\end{align*}
\]
For quarterly GDP and the first three monthly variables (industrial production, real manufacturing trade and sales, and real personal income excluding transfer payments), the following assumptions hold. Note that $\gamma_{10}^q$ is constrained to 1 for the identification of the model.

$$\gamma_1^q(L) = \gamma_{10}^q \equiv 1$$

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

For the fourth monthly variable (non-farm payroll employment), a richer lag structure is assumed, in order to take into account that employment may be lagging, as in Stock and Watson (1989):

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

We follow Chauvet and Hamilton (2005) in specifying an AR(1) dynamics for the idiosyncratic shocks and the underlying factor:

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

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

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

Lastly, it is assumed that 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 purely cyclical, with cyclical fluctuations captured by the underlying factor $\Delta c_t$ and the idiosyncratic shocks $u_{j,t}^m$. This implies $a_{j,t}^m = 0$ for $j = 1 \ldots 4$. We thus limit the parametrisation of our model and only explicitly account for low-frequency fluctuations of our main variable of interest, which is GDP growth. Antolin-Díaz 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 unit roots.

Additional details on the state-space representation of the model are available in Appendix 4.
4. Data and estimation strategy

4a. Data sources

In our information set, we consider five variables: quarterly real GDP and the four monthly variables used by the Business Cycle Dating Committee of the NBER, namely: industrial production, real manufacturing trade and sales, real personal income excluding transfer payments and non-farm payroll employment. The U.S. 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). 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-farm payroll 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). In the following, only real-time samples, corresponding to monthly releases of the FRED-MD database, are used.

4b. 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 $S_{1,T}$ and $V_{1,T}$ 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 $S_{1,T}$ and $V_{1,T}$. This is a key advantage as compared to the strategy put forward by Camacho et al. (2014) 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.

---

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

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:

- Draw the state vector $\alpha_{1,T}$ conditional on $\Delta y_{1,T}^\ast$, $S_{1,T}$, $V_{1,T}$ 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). This simulation smoother efficiently deals 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.

- Draw $S_{1,T}$ conditional on $\alpha_{1,T}$, $V_{1,T}$ and model parameters, based on Hamilton’s (1989) filter.

- Draw $V_{1,T}$ conditional on $\alpha_{1,T}$, $S_{1,T}$ and model parameters, based on Hamilton’s (1989) filter.

- Sequentially draw the different model parameters, based on $\Delta y_{1,T}^\ast$, $\alpha_{1,T}$, $S_{1,T}$, $V_{1,T}$ and the other model parameters.

Additional details on the estimation strategy are available in Appendices 3 and 4.
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.

5a. 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 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: Smoothed probability of being in a high-volatility regime stemming from the Extended MS-DFM

5b. 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 between 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 U.S.. 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

<table>
<thead>
<tr>
<th>Regime</th>
<th>Intercept and 95% probability interval</th>
<th>Implied expectation, given $\phi_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-volatility recession</td>
<td>- 0.69 [- 0.99 ; - 0.30]</td>
<td>- 0.87</td>
</tr>
<tr>
<td>Low-volatility recession</td>
<td>- 0.23 [- 0.41 ; - 0.06]</td>
<td>- 0.29</td>
</tr>
<tr>
<td>Low-volatility expansion</td>
<td>0.03 [- 0.00 ; 0.06]</td>
<td>0.04</td>
</tr>
<tr>
<td>High-volatility expansion</td>
<td>0.23 [0.03 ; 0.45]</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Figure 2: Influence of regime switches in volatility on smoothed recession probabilities

Figure 3: Smoothed and filtered probabilities of being in recession (extended MS-DFM)

5c. Declining long-term GDP growth in the U.S. since the turn of the century

Similarly to Antolin-Diaz et al. (2016), we identify a decrease in U.S. long-term GDP growth which started in the early 2000s, i.e. before the Great Recession. As compared to 2000, the U.S. long-term GDP growth estimate is lower by around one percentage point per year in 2010 and seems to have stabilised since then. This slowdown in U.S. long-term GDP growth follows an increase in the second half of the 1990s, which corresponds to a period of increase in U.S. productivity growth (Jorgenson et al. 2008).

Figure 4: Smoothed estimate of the U.S. annual long-run GDP growth rate, with 68% probability interval

---

5d. Comparison of different model specifications based on Deviance Information Criteria and marginal log-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 a statistical information criterion. More precisely, we 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 penalizing for model complexity:

\[
DIC = E_{y^T}[-2 \cdot \log f(\Delta y^T_{1:T}|\bar{\theta}, S^*_1, V^*_1) ]
\]

\[
\text{Decrease when model fit improves}
\]

\[
+ E_{y^T}[-2 \cdot \log f(\Delta y^T_{1:T}|\bar{\theta}, S^*_1, V^*_1) + 2 \log f(\Delta y^T_{1:T}|\bar{\theta}^*, S^*_1, V^*_1) ]
\]

\[
\text{Penalty for model complexity}
\]

In the above formula, \( \bar{\theta} \) corresponds to the vector of constant model parameters (e.g. autoregressive parameters), \( \bar{\theta}^* \) to their posterior mean, and \( f(\Delta y^T_{1:T}|\bar{\theta}, S^*_1, V^*_1) \) 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 conditional expectation and variance of the model’s state vector, conditional on past observations and other model variables and parameters, which then allows to estimate \( \log f(\Delta y^T_{1:T-1}|\bar{\theta}^*, S^*_1, V^*_1) \) for each date \( t \). In this case, the conditioning set includes the two Markov variables \( S^*_1 \) and \( V^*_1 \) and all constant model parameters, but not the large state vector \( \bar{\alpha}_{1:T} \). Summing these log-predictive likelihoods over time and averaging over posterior draws of the Gibbs Sampler finally gives an estimate of \( E_{y^T}[-2 \cdot \log f(\Delta y^T_{1:T-1}|\bar{\theta}^*, S^*_1, V^*_1) ] \). Alternatively, first averaging over posterior draws of the Gibbs Sampler, and then computing the predictive log-likelihoods with the Kalman filter allows to estimate \( \log f(\Delta y^T_{1:T}|\bar{\theta}^*, S^*_1, V^*_1) \).

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. In a nutshell, we use the Kalman filter to obtain \( f(\Delta y^T_{t}|\Delta y^T_{1:t-1}, \bar{\theta}^*, S^*_1, V^*_1) \) for each date \( t \) and then numerically integrate out \( S^*_1 \) and \( V^*_1 \) using a particle filter. This gives us \( f(\Delta y^T_{t}|\Delta y^T_{1:t-1}, \bar{\theta}^*) \) for each date,

\[^9\text{Note that given the structure of the model, the following equality applies: } f(\Delta y^T_{t}|\Delta y^T_{1:t-1}, \bar{\theta}, S^*_1, V^*_1) = f(\Delta y^T_{t}|\Delta y^T_{1:t-1}, \bar{\theta}, S^*_1, V^*_1).\]
from which we can compute $\log f(\Delta y_{1,T} | \tilde{\theta}^*)$. We finally rely on Chib’s (1995) formula to estimate the marginal log-likelihood $\log f(\Delta y_{1,T})$ of each specification, as follows:

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

As above, $\tilde{\theta}^*$ corresponds to the vector of constant model parameters evaluated at their posterior mean. $\log \varphi(\tilde{\theta}^*)$ and $\log \varphi(\tilde{\theta}^* | \Delta y_{1,T})$ are the log-prior and posterior densities of these parameters, respectively. These computations and the functioning of the particle filter are explained in detail in Appendix 6.

To the best of our knowledge, Kaufmann (2000) is the only other paper in the literature computing the marginal likelihood of MS-DFMs. Since this other paper considers a different information set (quarterly GDP, consumption and investment), a shorter sample ending in the mid-1990s, and models without switching volatility and time-varying growth rates, we consider the following results and the marginal likelihood computation method described in Appendix 6 as contributions of the present paper.

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. Allowing for time-variation in the long-term GDP growth rate further reduces the DIC, although less significantly. Similarly, the main improvement to the marginal log-likelihood is due to the inclusion of the Markov-Switching volatility into the model. Given the closeness of several marginal log-likelihoods, we conclude that these statistics allow distinguishing two groups of models: the linear DFM and the MS-DFM with a Markov-Switching intercept on the one hand, the MS-DFM with Markov-Switching intercept and volatility and the extended MS-DFM on the other hand. The second group of model is the preferred one.

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 DIC and 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

<table>
<thead>
<tr>
<th>Model specification</th>
<th>DIC</th>
<th>Marginal log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear DFM</td>
<td>3046.1</td>
<td>-1534.6</td>
</tr>
<tr>
<td>MS-DFM with MS on intercept only</td>
<td>2990.5</td>
<td>-1534.0</td>
</tr>
<tr>
<td>MS-DFM with MS on intercept and volatility</td>
<td>2867.3</td>
<td>-1495.9</td>
</tr>
<tr>
<td>MS-DFM with MS on intercept and volatility, and time-varying long-term GDP growth rate (i.e. extended MS-DFM)</td>
<td>2843.4</td>
<td>-1494.6</td>
</tr>
</tbody>
</table>

Note: In all cases, the posterior mean of model parameters is estimated based on 15000 draws of the Gibbs Sampler, from which the first 5000 are discarded. In addition, 500 particles are used for marginal log-likelihood computations (see Appendix 6 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.


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 a new monthly data point an re-estimate the model. However, the specification of the model remains the same.

6a. Real-time estimation of the U.S. long-term GDP growth rate

As already pointed out in the previous section, the decrease in the U.S. long-term GDP growth rate that we estimate with our extended model is about 1 percentage point per year compared to the early 2000s (see red curve in Figure 5, 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.¹⁰

Figure 5 shows that the model needs some time to identify the drop in long-run GDP growth. Until 2010, the estimated long-run GDP growth is above 3%, reaching 4% at some occasions, but it progressively decreases from 2011 onwards (dashed curves in Figure 5). It appears that the model-based real-time

¹⁰ 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.
estimates are closer to the final estimate (red curve) than the SPF real-time estimates since 2011, while the converse holds true from 2007 to 2010.

Figure 5: Smoothed estimates of the U.S. annual long-run GDP growth rate, consecutive data vintages (2007M01 - 2017M01)
6b. Real-time detection of the Great Recession

Whereas we relied on the 2017M12 data vintage in order to estimate filtered and smoothed 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). 11 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 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 7 and 8 months earlier than the NBER, respectively, thus allowing to integrate this information into GDP forecasts in a more timely way.

<table>
<thead>
<tr>
<th>Peak date - NBER</th>
<th>Peak date - Extended MS-DFM</th>
<th>Peak date announcement - NBER</th>
<th>Peak date announcement – Extended MS-DFM</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007 M12 (Start of the Great Recession)</td>
<td>2007 M09</td>
<td>2008 M12</td>
<td>2008 M05 (-7m)</td>
</tr>
<tr>
<td>Trough date - NBER</td>
<td>Trough date - Extended MS-DFM</td>
<td>Trough date announcement - NBER</td>
<td>Trough date announcement – Extended MS-DFM</td>
</tr>
<tr>
<td>2009 M06 (End of the Great Recession)</td>
<td>2009 M06</td>
<td>2010 M09</td>
<td>2010 M01 (-8m)</td>
</tr>
</tbody>
</table>

11 Hamilton (2011) suggests the following decision rule when relying on quarterly GDP data: “When the one-quarter smoothed inference $P(S_t | y_{t+1}, y_{t+2}, \ldots, y_1; \hat{\theta}_{t+1} )$ 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 $P(S_{t-j} | y_{t+1}, y_{t+2}, \ldots, y_1; \hat{\theta}_{t+1} )$ 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.
6c. Real-time forecasting performance of U.S. 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 U.S. 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 6.

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 6: Calendar of the forecasting exercise

Table 5 compares the real-time Root Mean Square Forecast Errors (RMSFEs) of our extended MS-DFM vis-à-vis those stemming from a standard linear DFM. We present the ratios of RMFES, meaning that a value below one reveals that the extended MS-DFM outperform the DFM. Forecast horizons are taken from \(h=6\) months to \(h=1\) month before the GDP release date. 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. Figures 7 and 8 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 by 5 to 10% at
horizons (M-4) to (M-6) in both subperiods. 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 back 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 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 from four to six months before the GDP release date. Figure 7 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 transition 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)

<table>
<thead>
<tr>
<th></th>
<th>(M-1)</th>
<th>(M-2)</th>
<th>(M-3)</th>
<th>(M-4)</th>
<th>(M-5)</th>
<th>(M-6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007Q4-2009Q2</td>
<td>1.13</td>
<td>1.19</td>
<td>1.06</td>
<td>0.92</td>
<td>0.99</td>
<td>0.94</td>
</tr>
<tr>
<td>2011Q1-2017Q4</td>
<td>0.90</td>
<td>0.86</td>
<td>0.87</td>
<td>0.91</td>
<td>0.88</td>
<td>0.87</td>
</tr>
</tbody>
</table>
Figure 7: Comparison of point forecasts at a 6-month horizon: linear DFM vs. Extended MS-DFM specification

![Graph showing comparison of point forecasts at a 6-month horizon.](image)

Figure 8: Comparison of point forecasts at a 1-month horizon: linear DFM vs. Extended MS-DFM specification

![Graph showing comparison of point forecasts at a 1-month horizon.](image)
7. Conclusion

In this paper, we put forward an extension of the standard Markov-Switching Dynamic Factor Model (MS-DFM), by accounting for two stylized facts of the U.S. economy, namely switches in macroeconomic volatility and a long-run declining trend in GDP. This extended MS-DFM comes as an innovation in the literature and could be potentially used for various advanced economies showing similar patterns.

More specifically, we show that the introduction of a Markov-Switching volatility in the standard MS-DFM framework is supported by statistical criteria (Deviation Information Criterion and marginal likelihoods) and largely improves the detection of turning points in the U.S. business cycle during the Great Moderation period. In addition, our extended model shows evidence that the Great Moderation period that started in the eighties is not over, as the Great Recession can be considered as a temporary increase in macroeconomic volatility. This empirical result is extremely useful for the detection of future recessions, as our model is able to detect recessions in both high- and low-volatility macroeconomic environments.

As a second improvement to the standard MS-DFM framework, our extended model also incorporates a time-varying long-term GDP growth rate. This feature enables to identify a continuous decline in the U.S. long-term GDP growth rate of around one percentage point per year as compared to the early 2000s. This result is line with the empirical literature on the long-run decline in U.S. potential growth. According to the model, about half of this loss would be related to the Great Recession.

Overall, our new extended model empirically shows that allowing for both regime switches and time-variation in long-term GDP growth is important to improve the accuracy of real-time GDP short-term 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.
References


APPENDIX 1: Input data and underlying factor

In this Appendix, 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 A1.1: Real GDP\textsuperscript{12}

\textsuperscript{12} Quarterly GDP growth rates are reported in the third month of each quarter (square marks).
Figure A1.2: Industrial production

Figure A1.3: Real personal income excluding transfer payments
Figure A1.4: Real manufacturing trade and sales

Figure A1.5: Non-farm payroll employment
Figure A1.6: Underlying factor and NBER recessions

Figure A1.7: Underlying factor and smoothed probability of being in a high-volatility regime

The comparison with Chauvet and Piger (2008) is based on the updated smoothed recession probabilities that Jeremy Piger publishes every month on its website\textsuperscript{13}. We rely on the vintage that he published in December 2017, which is also our last vintage of data. Figure A6.1 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 A6.1: Comparison of the smoothed recession probabilities updated by J. Piger and those obtained with the extended MS-DFM

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure.png}
\caption{Comparison of the smoothed recession probabilities updated by J. Piger and those obtained with the extended MS-DFM}
\end{figure}


\textsuperscript{13} See \url{http://pages.uoregon.edu/jpiger/us_recession_probs.htm/history_of_real_time_recess.xls}
APPENDIX 3: Prior and posterior parameter distributions of the extended MS-DFM

Table A2.1 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 2 in order to limit the influence of the priors. See Appendix 4 for additional details on the calibration of the priors.

Table A2.1: Prior and posterior parameter distributions of the extended MS-DFM

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Prior</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Posterior</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>95% bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_{10}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>1.99</td>
<td>0.13</td>
<td>1.75</td>
<td>2.24</td>
</tr>
<tr>
<td>$y_{20}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>1.83</td>
<td>0.13</td>
<td>1.57</td>
<td>2.09</td>
</tr>
<tr>
<td>$y_{30}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>0.78</td>
<td>0.09</td>
<td>0.61</td>
<td>0.97</td>
</tr>
<tr>
<td>$y_{40}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>0.58</td>
<td>0.05</td>
<td>0.50</td>
<td>0.68</td>
</tr>
<tr>
<td>$y_{50}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>-0.11</td>
<td>0.04</td>
<td>-0.19</td>
<td>-0.04</td>
</tr>
<tr>
<td>$y_{60}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>0.13</td>
<td>0.03</td>
<td>0.07</td>
<td>0.20</td>
</tr>
<tr>
<td>$y_{70}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>0.10</td>
<td>0.03</td>
<td>0.05</td>
<td>0.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Prior</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Posterior</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>95% bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\sigma_{1}^{2})^{2}$</td>
<td>IG(2, $\frac{9}{38}$)</td>
<td>0.15</td>
<td>--</td>
<td>0.24</td>
<td>0.04</td>
<td>0.15</td>
<td>0.31</td>
</tr>
<tr>
<td>$(\sigma_{2}^{m})^{2}$</td>
<td>IG(2, Var($\Delta y_{10}$))</td>
<td>0.54</td>
<td>--</td>
<td>0.23</td>
<td>0.02</td>
<td>0.19</td>
<td>0.26</td>
</tr>
<tr>
<td>$(\sigma_{3}^{m})^{2}$</td>
<td>IG(2, Var($\Delta y_{20}$))</td>
<td>0.93</td>
<td>--</td>
<td>0.56</td>
<td>0.04</td>
<td>0.49</td>
<td>0.64</td>
</tr>
<tr>
<td>$(\sigma_{4}^{m})^{2}$</td>
<td>IG(2, Var($\Delta y_{30}$))</td>
<td>0.38</td>
<td>--</td>
<td>0.33</td>
<td>0.02</td>
<td>0.29</td>
<td>0.37</td>
</tr>
<tr>
<td>$(\sigma_{5}^{m})^{2}$</td>
<td>IG(2, Var($\Delta y_{40}$))</td>
<td>0.04</td>
<td>--</td>
<td>4.79e-3</td>
<td>0.99e-3</td>
<td>0.30e-3</td>
<td>0.70e-3</td>
</tr>
<tr>
<td>$\phi_{1}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>-0.2</td>
<td>0.05</td>
<td>-0.30</td>
<td>-0.10</td>
</tr>
<tr>
<td>$\phi_{2}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>-0.40</td>
<td>0.04</td>
<td>-0.47</td>
<td>-0.31</td>
</tr>
<tr>
<td>$\phi_{3}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>-0.16</td>
<td>0.04</td>
<td>-0.24</td>
<td>-0.08</td>
</tr>
<tr>
<td>$\phi_{4}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>0.70</td>
<td>0.07</td>
<td>0.54</td>
<td>0.83</td>
</tr>
<tr>
<td>$\phi_{5}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>0.20</td>
<td>0.16</td>
<td>-0.04</td>
<td>0.58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Prior</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Posterior</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>95% bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\sigma_{6,1})^{2}$</td>
<td>IG(2, 5e-3)</td>
<td>5.00e-3</td>
<td>--</td>
<td>8.93e-5</td>
<td>5.76e-5</td>
<td>7.42e-5</td>
<td>1.23e-4</td>
</tr>
<tr>
<td>$PrS_{00}$</td>
<td>Beta(90, 7)</td>
<td>0.93</td>
<td>2.61e-2</td>
<td>0.91</td>
<td>2.34e-2</td>
<td>0.86</td>
<td>0.95</td>
</tr>
<tr>
<td>$PrS_{01}$</td>
<td>Beta(470, 7)</td>
<td>0.99</td>
<td>5.50e-3</td>
<td>0.98</td>
<td>4.79e-3</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td>$PrV_{00}$</td>
<td>Beta(285, 1)</td>
<td>0.99</td>
<td>2.58e-3</td>
<td>0.99</td>
<td>4.84e-3</td>
<td>0.98</td>
<td>1.00</td>
</tr>
<tr>
<td>$PrV_{01}$</td>
<td>Beta(285, 1)</td>
<td>0.99</td>
<td>2.58e-3</td>
<td>0.98</td>
<td>9.19e-3</td>
<td>0.97</td>
<td>1.00</td>
</tr>
<tr>
<td>$h$</td>
<td>IG(5, 10)</td>
<td>2.50</td>
<td>1.44</td>
<td>1.15</td>
<td>0.75</td>
<td>0.05</td>
<td>2.84</td>
</tr>
<tr>
<td>$\mu_{00}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>-0.23</td>
<td>0.08</td>
<td>-0.41</td>
<td>-0.06</td>
</tr>
<tr>
<td>$\mu_{01}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>0.26</td>
<td>0.08</td>
<td>0.09</td>
<td>0.42</td>
</tr>
<tr>
<td>$\mu_{10}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>-0.46</td>
<td>0.13</td>
<td>-0.68</td>
<td>-0.16</td>
</tr>
<tr>
<td>$\mu_{11}$</td>
<td>N(0, 1)</td>
<td>0</td>
<td>1</td>
<td>0.66</td>
<td>0.20</td>
<td>0.27</td>
<td>0.99</td>
</tr>
</tbody>
</table>


14 Remind that $y_{10}^q$ is set equal to 1 for identification purposes, and not estimated (see Section 2 and Appendix 4).
APPENDIX 4: State-space representation of the model and description of the estimation methodology

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:

\[
\begin{align*}
\Delta y^q_{it} &= \left(\frac{1}{3} a^q_{1t} + \frac{2}{3} a^q_{1t-1} + a^q_{1t-2} + \frac{2}{3} a^q_{1t-3} + \frac{1}{3} a^q_{1t-4}\right) + \gamma^q_{10} \left(\frac{1}{3} \Delta c_t + \frac{2}{3} \Delta c_{t-1} + \Delta c_{t-2} + \frac{2}{3} \Delta c_{t-3} + \frac{1}{3} \Delta c_{t-4}\right) \\
&\quad + \left(\frac{1}{3} a^q_{1t} + \frac{2}{3} u^q_{1t-1} + u^q_{1t-2} + \frac{2}{3} u^q_{1t-3} + \frac{1}{3} u^q_{1t-4}\right) \\
\Delta y^m_{jt} &= y^m_j(L) \Delta c_t + u^m_{jt} \text{ for } j = 1 \ldots 4 \\
a^q_{1t} &= a^q_{1t-1} + \sigma_a \cdot \eta^a_t ; \quad \eta^a_t \sim N(0,1) \\
\phi(L) \Delta c_t &= \mu_{c_t} + \sqrt{1+h} \cdot V_t \cdot \sigma_c \cdot \eta^c_t ; \quad \eta^c_t \sim N(0,1) \\
\psi^q_1(L) u^q_{1t} &= u^q_{1t} - \psi^q_{11} u^q_{1t-1} = \sigma^q_1 \cdot \epsilon^q_{1t} ; \quad \epsilon^q_{1t} \sim N(0,1) \\
\psi^m_j(L) u^m_{jt} &= u^m_{jt} - \psi^m_{11} u^m_{jt-1} = \sigma^m_j \cdot \epsilon^m_{jt} ; \quad \epsilon^m_{jt} \sim N(0,1) \text{ for } j = 1 \ldots 4
\end{align*}
\]

Representation 2:

\[
\begin{align*}
\Delta y^q_{it} &= \left(\frac{1}{3} a^q_{1t} + \frac{2}{3} a^q_{1t-1} + a^q_{1t-2} + \frac{2}{3} a^q_{1t-3} + \frac{1}{3} a^q_{1t-4}\right) + \gamma^q_{10} \left(\frac{1}{3} \Delta c_t + \frac{2}{3} \Delta c_{t-1} + \Delta c_{t-2} + \frac{2}{3} \Delta c_{t-3} + \frac{1}{3} \Delta c_{t-4}\right) \\
&\quad + \left(\frac{1}{3} a^q_{1t} + \frac{2}{3} u^q_{1t-1} + u^q_{1t-2} + \frac{2}{3} u^q_{1t-3} + \frac{1}{3} u^q_{1t-4}\right) \\
\psi^m_j(L) \Delta y^m_{jt} &= \Delta y^m_{jt} \equiv \gamma^m_j(L) \psi^m_j(L) \Delta c_t + \sigma^m_j \cdot \epsilon^m_{jt} = y^m_j(L) \psi^m_j(L) \Delta c_t + \sigma^m_j \cdot \epsilon^m_{jt} ; \quad \epsilon^m_{jt} \sim N(0,1) \text{ for } j = 1 \ldots 4 \\
a^q_{1t} &= a^q_{1t-1} + \sigma_a \cdot \eta^a_t ; \quad \eta^a_t \sim N(0,1) \\
\phi(L) \Delta c_t &= \mu_{c_t} + \sqrt{1+h} \cdot V_t \cdot \sigma_c \cdot \eta^c_t ; \quad \eta^c_t \sim N(0,1) \\
\psi^q_1(L) u^q_{1t} &= u^q_{1t} - \psi^q_{11} u^q_{1t-1} = \sigma^q_1 \cdot \epsilon^q_{1t} ; \quad \epsilon^q_{1t} \sim N(0,1)
\end{align*}
\]

Using representation 2, the model can be cast in state-space from, as follows:
\[
\begin{align*}
\begin{pmatrix}
\Delta y_{1t}^q \\
\Delta y_{2t}^m \\
\Delta y_{3t}^m \\
\Delta y_{4t}^m
\end{pmatrix} &= 
\begin{pmatrix}
\psi_1^m(L) \Delta y_{1t}^m \\
\psi_2^m(L) \Delta y_{2t}^m \\
\psi_3^m(L) \Delta y_{3t}^m \\
\psi_4^m(L) \Delta y_{4t}^m
\end{pmatrix} = 
\begin{pmatrix}
\frac{2}{3} \gamma_{10}^q \\
\frac{3}{3} \gamma_{11}^q \\
\frac{3}{3} \gamma_{12}^q \\
\frac{3}{3} \gamma_{13}^q \\
\frac{3}{3} \gamma_{14}^q
\end{pmatrix} \begin{pmatrix}
\gamma_{10}^q \\
\gamma_{11}^q \\
\gamma_{12}^q \\
\gamma_{13}^q \\
\gamma_{14}^q
\end{pmatrix} + 
\begin{pmatrix}
\Delta c_t \\
\Delta c_{t-1} \\
\Delta c_{t-2} \\
\Delta c_{t-3} \\
\Delta c_{t-4}
\end{pmatrix} + 
\begin{pmatrix}
\mu_{S_t V_t} \\
\mu_{S_{t-1} V_{t-1}} \\
\mu_{S_{t-2} V_{t-2}} \\
\mu_{S_{t-3} V_{t-3}} \\
\mu_{S_{t-4} V_{t-4}}
\end{pmatrix} + 
\begin{pmatrix}
\sqrt{1 + h \cdot V_t \cdot \sigma_c}, \eta_t^q \\
\sqrt{1 + h \cdot V_t \cdot \sigma_c}, \eta_t^q \\
\sqrt{1 + h \cdot V_t \cdot \sigma_c}, \eta_t^q \\
\sqrt{1 + h \cdot V_t \cdot \sigma_c}, \eta_t^q \\
\sqrt{1 + h \cdot V_t \cdot \sigma_c}, \eta_t^q
\end{pmatrix}.
\end{align*}
\]
Notations:
\[ y_j^m (L) \equiv y_j^m (L) \psi_j^m (L) = \left( \gamma_{j0}^m + \gamma_{j1}^m L + \gamma_{j2}^m L^2 + \gamma_{j3}^m L^3 \right)(1 - \psi_{j1}^m L) \]
\[ = \gamma_{j0}^m + \left( -\gamma_{j0}^m \psi_{j1}^m + \gamma_{j1}^m \right) L + \left( -\gamma_{j1}^m \psi_{j1}^m + \gamma_{j2}^m \right) L^2 + \left( -\gamma_{j2}^m \psi_{j1}^m + \gamma_{j3}^m \right) L^3 + \left( -\gamma_{j3}^m \psi_{j1}^m \right) L^4 \]

With \( y_{j1}^m = y_{j2}^m = y_{j3}^m = 0 \) for \( j = 1,3 \), the following notations hold:
\[
\begin{align*}
\gamma_{j0}^m &\equiv \gamma_{j0}^m \\
\gamma_{j1}^m &\equiv -\gamma_{j1}^m \psi_{j1}^m \\
\gamma_{j2}^m &\equiv -\gamma_{j2}^m \psi_{j1}^m \\
\gamma_{j3}^m &\equiv -\gamma_{j3}^m \psi_{j1}^m \\
\gamma_{41}^m &\equiv -\gamma_{40}^m \psi_{41}^m + \gamma_{41}^m \\
\gamma_{42}^m &\equiv -\gamma_{41}^m \psi_{41}^m + \gamma_{42}^m \\
\gamma_{43}^m &\equiv -\gamma_{42}^m \psi_{41}^m + \gamma_{43}^m \\
\gamma_{44}^m &\equiv -\gamma_{43}^m \psi_{41}^m 
\end{align*}
\]

Identification of the model:
Kim and Nelson (1998) consider a model without volatility switches, nor low-frequency fluctuations:
\[
\begin{align*}
\Delta y_{it} &= y_i(L) \Delta c_t + u_{it} \\
\psi_i(L) u_{it} &= \sigma_c \cdot \epsilon_{it} ; \quad \epsilon_{it} \sim \text{IID } N(0,1) \\
\phi(L)(\Delta c_t - \mu_{c_t}) &= \sigma_c \cdot \eta_t ; \quad \eta_t \sim N(0,1)
\end{align*}
\]
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

\[ p(\bar{a}_{1:T}, S_{1:T}, V_{1:T}, \bar{y}, \bar{\psi}, a_q^m, c, \phi_1, \bar{\mu}, h, \alpha_c, a, PrS_{00}, PrS_{11}, PrV_{00}, PrV_{11}|\Delta y_{1:T}) \].

The algorithm sequentially draws from the following conditional distributions:

\[ p(\bar{a}_{1:T}|\Delta y_{1:T}, S_{1:T}, V_{1:T}, \bar{\theta}) \]

\[ p(S_{1:T}|\Delta y_{1:T}, \bar{a}_{1:T}, V_{1:T}, \bar{\theta}) = p(S_{1:T}|\Delta \bar{y}_{1:T}, \Delta \bar{c}_{1:T}, V_{1:T}, \bar{\theta}) \]

\[ p(V_{1:T}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, S_{1:T}, \bar{\theta}) = p(V_{1:T}|\Delta \bar{y}_{1:T}, \bar{c}_{1:T}, S_{1:T}, \bar{\theta}) \]

\[ p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, S_{1:T}, V_{1:T}, \bar{\theta}|\bar{\psi}) = p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, \bar{\psi}, \bar{\sigma}^m) \]

\[ p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, S_{1:T}, V_{1:T}, \bar{\theta}|\bar{\psi}) = p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, \bar{\psi}, \bar{\sigma}^m, a_q^r) \]

\[ p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, S_{1:T}, V_{1:T}, \bar{\theta}|\bar{\sigma}^q, \bar{\sigma}^m) = p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, \bar{\psi}, \bar{\sigma}^q, \bar{\sigma}^m) \]

\[ p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, S_{1:T}, V_{1:T}, \bar{\theta}|\bar{\psi}) = p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, \bar{\psi}, \bar{\sigma}^m) \]

\[ p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, S_{1:T}, V_{1:T}, \bar{\theta}|\bar{\psi}) = p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, \bar{\psi}, \bar{\sigma}^m, a_q^r) \]

\[ p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, S_{1:T}, V_{1:T}, \bar{\theta}|\bar{\psi}) = p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, \bar{\psi}, \bar{\sigma}^m) \]

\[ p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, S_{1:T}, V_{1:T}, \bar{\theta}|\bar{\psi}) = p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, \bar{\psi}, \bar{\sigma}^m, a_q^r) \]

\[ p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, S_{1:T}, V_{1:T}, \bar{\theta}|\bar{\psi}) = p(\bar{y}|\Delta \bar{y}_{1:T}, \bar{a}_{1:T}, \bar{\psi}, \bar{\sigma}^m) \]

where the following notations are used:

\[ \Delta y_{1:T} \equiv (\Delta y_1, ..., \Delta y_T); \bar{a}_{1:T} \equiv (\alpha_1, ..., \alpha_T); S_{1:T} \equiv (S_1, ..., S_T); V_{1:T} \equiv (V_1, ..., V_T) \]

\[ \bar{\gamma} \equiv (\gamma_{1,0}^q, \gamma_{1,0}^m, \gamma_{2,0}^m, \gamma_{3,0}^m, \gamma_{4,0}^m, \gamma_{4,1}^m, \gamma_{4,2}^m, \gamma_{4,3}^m) \]

\[ \bar{\psi} \equiv (\psi_{11}^q, \psi_{11}^m, \psi_{21}^m, \psi_{31}^m, \psi_{41}^m) \]

\[ \bar{\sigma}^m \equiv (a_1^m, a_2^m, a_3^m, a_4^m) \]

\[ \bar{\mu} \equiv (\mu_{00}, \mu_{01}, \mu_{10}, \mu_{11}) \]

\[ \bar{\theta} \equiv (\bar{y}, \bar{\psi}, \bar{\sigma}^m, \phi_1, \bar{\mu}, h, \alpha_c, a, PrS_{00}, PrS_{11}, PrV_{00}, PrV_{11}) \]
**Step 1:** Based on Representation 2, draw $\alpha_{1:T}$ conditional on $\Delta \ddot{y}_{1:T}^r$, $\Delta \ddot{c}_{1:T}$, $\ddot{V}_{1:T}$ 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 $\hat{S}_{1:T}$ conditional on $\Delta \ddot{y}_{1:T}^r$, $\Delta \ddot{c}_{1:T}$, $\ddot{V}_{1:T}$ and model parameters\(^{15}\):

$$p(S_{1:T} | \Delta \ddot{y}_{1:T}^r, \Delta \ddot{c}_{1:T}, \ddot{V}_{1:T}, \theta) = p(S_T | \Delta \ddot{c}_{1:T}, \ddot{V}_{1:T}, \theta) \cdot \prod_{t=1}^{T-1} p(S_t | S_{t+1}, \Delta \ddot{c}_{1:t}, \ddot{V}_{1:T}, \theta)$$

**Step 2a:** Run Hamilton’s (1989) filter to get $p(S_T | \Delta \ddot{c}_{1:T}, \ddot{V}_{1:T}, \theta)$ and $p(S_t | \Delta \ddot{c}_{1:t}, \ddot{V}_{1:T}, \theta)$ for $t = 1..T-1$

**Step 2a1:** Initialise the filter with unconditional probabilities

$$p(S_0 = j | \Delta c_0, \ddot{V}_{1:T}, \theta) = p(S_0 = j)$$

**Step 2a2:** from $p(S_{t-1} = j | \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta)$ to $p(S_t = k | \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta)$

$$p(S_t = k | \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta) = \sum_j p(S_t = k, S_{t-1} = j | \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta)$$

$$= \sum_j p(S_t = k | S_{t-1} = j, \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta) \cdot p(S_{t-1} = j | \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta)$$

**Step 2a3:** from $p(S_t = k | \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta)$ to $p(S_t = k | \Delta \ddot{c}_{1:t}, \ddot{V}_{1:T}, \theta)$

$$p(S_t = k | \Delta \ddot{c}_{1:t}, \ddot{V}_{1:T}, \theta) = \frac{f(\Delta c_t | S_t = k, \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta) \cdot p(S_t = k | \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta)}{f(\Delta c_t | \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta)}$$

$$f(\Delta c_t | \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta) = \sum_k f(\Delta c_t | S_t = k, \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta) \cdot p(S_t = k | \Delta \ddot{c}_{1:t-1}, \ddot{V}_{1:T}, \theta)$$

**Step 2a4:** Store $p(S_t = k | \Delta \ddot{c}_{1:t}, \ddot{V}_{1:T}, \theta)$ as an intermediate output of Hamilton’s filter.

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

**Step 2b:** from $p(S_t | \Delta \ddot{c}_{1:t}, \ddot{V}_{1:T}, \theta)$ to $p(S_t | S_{t+1}, \Delta \ddot{c}_{1:t}, \ddot{V}_{1:T}, \theta)$

$$p(S_t | S_{t+1}, \Delta \ddot{c}_{1:t}, \ddot{V}_{1:T}, \theta) = \frac{p(S_{t+1} | S_{t}, \Delta \ddot{c}_{1:t}, \ddot{V}_{1:T}, \theta) \cdot p(S_t | \Delta \ddot{c}_{1:t}, \ddot{V}_{1:T}, \theta)}{p(S_{t+1} | \Delta \ddot{c}_{1:t}, \ddot{V}_{1:T}, \theta)}$$

\(^{15}\) See proof in Appendix 5.
\[
= \frac{p(S_{t+1}|S_t) \cdot p(S_t|\Delta c_{1,T}, V_{1,T}, \theta) \cdot p(\Delta c_{1,T})}{\sum_j p(S_{t+1}|S_t = j \cdot \Delta c_{1,T}, V_{1,T}, \theta) \cdot p(S_t = j \cdot \Delta c_{1,T}, V_{1,T}, \theta)}
\]

\(p(S_t|\Delta c_{1,T}, V_{1,T}, \theta)\) 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 \(\tilde{V}_{1,T}\) conditional on \(\Delta y_{1,T}, \Delta c_{1,T}, S_{1,T}\) and model parameters\(^{16}\):

\[
p(V_{t-1}^-|\Delta y_{1,T}, \Delta c_{1,T}, S_{1,T}, \theta) = \sum_j p(V_t = k, V_{t-1} = j \cdot \Delta c_{1,T}, S_{1,T}, \theta) \\
= \sum_j p(V_t = k, V_{t-1} = j \cdot \Delta c_{1,T}, S_{1,T}, \theta) \cdot p(V_{t-1} = j \cdot \Delta c_{1,T}, S_{1,T}, \theta)
\]

**Step 3a**: Run Hamilton’s (1989) filter to get \(p(V_t|\Delta c_{1,T}, S_{1,T}, \theta)\) and \(p(V_t|\Delta c_{1,T}, S_{1,T}, \theta)\) for \(t = 1..T-1\).

**Step 3a1**: Initialise the filter with unconditional probabilities

\[
p(V_0 = j|\Delta c_{0,}, S_{1,T}, \theta) = p(V_0 = j)
\]

**Step 3a2**: from \(p(V_{t-1} = j \cdot \Delta c_{1,T-1}, S_{1,T}, \theta)\) to \(p(V_t = k \cdot \Delta c_{1,T-1}, S_{1,T}, \theta)\)

\[
p(V_t = k \cdot \Delta c_{1,T-1}, S_{1,T}, \theta) = \sum_j p(V_t = k, V_{t-1} = j \cdot \Delta c_{1,T-1}, S_{1,T}, \theta) \\
= \sum_j p(V_t = k, V_{t-1} = j \cdot \Delta c_{1,T-1}, S_{1,T}, \theta) \cdot p(V_{t-1} = j \cdot \Delta c_{1,T-1}, S_{1,T}, \theta)
\]

**Step 3a3**: from \(p(V_t = k \cdot \Delta c_{1,T-1}, S_{1,T}, \theta)\) to \(p(V_t = k \cdot \Delta c_{1,T}, S_{1,T}, \theta)\)

\[
p(V_t = k \cdot \Delta c_{1,T}, S_{1,T}, \theta) = \frac{f(\Delta c_t|V_t = k, \Delta c_{1,T-1}, S_{1,T}, \theta) \cdot p(V_t = k \cdot \Delta c_{1,T-1}, S_{1,T}, \theta)}{f(\Delta c_t|\Delta c_{1,T-1}, S_{1,T}, \theta)} \\
f(\Delta c_t|\Delta c_{1,T-1}, S_{1,T}, \theta) = \sum_k f(\Delta c_t|V_t = k, \Delta c_{1,T-1}, S_{1,T}, \theta) \cdot p(V_t = k \cdot \Delta c_{1,T-1}, S_{1,T}, \theta)
\]

**Step 3a4**: Store \(p(V_t = k \cdot \Delta c_{1,T}, S_{1,T}, \theta)\) as an intermediate output of Hamilton’s filter.

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

\(^{16}\) The proof is similar to the one given in Appendix 5.
Step 3b: from \( p(V_t | \Delta c_{1,T}, S_{1,T}, \theta) \) to \( p(V_t | V_{t+1}, \Delta c_{1,T}, S_{1,T}, \theta) \)

\[
p(V_t | V_{t+1}, \Delta c_{1,T}, S_{1,T}, \theta) = \frac{p(V_{t+1} | V_t, \Delta c_{1,T}, S_{1,T}, \theta) \cdot p(V_t | \Delta c_{1,T}, S_{1,T}, \theta)}{p(V_{t+1} | \Delta c_t, \theta)}
\]

\[
= \frac{p(V_{t+1} | V_t) \cdot p(V_t | \Delta c_{1,T}, S_{1,T}, \theta)}{\sum_j p(V_{t+1} | V_t = j, \Delta c_{1,T}, S_{1,T}, \theta) \cdot p(V_t = j | \Delta c_{1,T}, S_{1,T}, \theta)}
\]

\( p(V_t | \Delta c_{1,T}, S_{1,T}, \theta) \) 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.

Step 4: For \( j=1..4 \), draw \( y_j^m \) conditional on \( \Delta y_{j,1,T}^m, \Delta c_{j,1,T}^m \) and \( \sigma_j^m \), based on representation 2 and the 4 following independent equations:

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

Prior distributions:

\( y_{10}^m \sim N(y_j^m, A_j^m) \) for \( j = 1..3 \) and \( y_4^m \equiv \begin{pmatrix} y_{40}^m \\ y_{41}^m \\ y_{42}^m \\ y_{43}^m \end{pmatrix} \sim N(y_4^m, A_4^m) \)

Following Kim and Nelson (1998), we use \( y_{j1}^m = 0 \) and \( A_j^m = 1 \) for \( j = 1..3 \), \( y_4^m = 0_{4x1} \) and \( A_4^m = I_{d4} \).

Notation: \( \Delta c_{j,1,T}^m \) is the (T x N_j) vector of right-hand-side variables in the above equations.

Posterior distributions:

\[
N \left( \left[ (A_j^m)^{-1} + (\sigma_j^m)^{-2} \Delta c_{j,1,T}^m \Delta c_{j,1,T}^m \right]^{-1} \left[ (A_j^m)^{-1} y_{j1}^m + (\sigma_j^m)^{-2} \Delta c_{j,1,T}^m \Delta y_{j,1,T}^m \right] \right)
\]

Step 5a: Draw \( \psi_1^q \) conditional on \( u_{1,1,T}^q \) and \( \sigma_1^q \), based on the following equation:

\[
u_{1,t}^q = \psi_{11}^q u_{1,t-1}^q + \sigma_1^q \cdot \varepsilon_{1t}^q
\]
Prior distribution: \( \psi_1^q \sim N \left( \psi_1^q, B_1^q \right) \). Following Kim and Nelson (1998), we use \( \psi_1 = 0 \) and \( B_1^q = 1 \).

Notations: \( \bar{u}_{i,1,T} \equiv \begin{pmatrix} u_{i,1}^q \\ \vdots \\ u_{i,T}^q \end{pmatrix} \)

Posterior distribution:
\[
N \left( \left( B_1^q \right)^{-1} + \left( \sigma_1^q \right)^{-2} u_{1,1,T-1} \bar{u}_{1,1,T-1} \right)^{-1} \left[ \left( B_1^q \right)^{-1} \psi_1^q + \left( \sigma_1^q \right)^{-2} u_{1,1,T-1} \bar{u}_{1,1,T-1} \right], \\
\left[ \left( B_1^q \right)^{-1} + \left( \sigma_1^q \right)^{-2} u_{1,1,T-1} \bar{u}_{1,1,T-1} \right]^{-1} \right)
\]

**Step 5b:** For \( j = 1 \ldots 4 \), draw \( \psi_j^m \) conditional on \( \Delta y_{j,1,T}^m, \Delta c_{1,T}^m \) and \( \sigma_j^m \), based on representation 2 and the 4 following independent equations:
\[
\frac{\Delta y_{j,t}^m - y_j^m(L) \Delta c_t}{\equiv x_{j,t}^m} = \psi_j^m \left( \Delta y_{j,t-1}^m - y_j^m(L) \Delta c_{t-1} \right) + \sigma_j^m \cdot \varepsilon_{jt}
\]

Prior distributions: \( \psi_j^m \sim N \left( \psi_j^m, B_j^m \right) \). Following Kim and Nelson (1998), we use \( \psi_j^m = 0 \) and \( B_j^m = 1 \).

Notations\(^{17}\): \( x_{j,t}^m \equiv \Delta y_{j,t}^m - y_j^m(L) \Delta c_t \); \( \bar{x}_{j,1,T}^m \equiv \begin{pmatrix} \Delta y_{j,1}^m - y_j^m(L) \Delta c_1 \\ \vdots \\ \Delta y_{j,T}^m - y_j^m(L) \Delta c_T \end{pmatrix} \)

Posterior distributions:
\[
N \left( \left( B_j^m \right)^{-1} + \left( \sigma_j^m \right)^{-2} \bar{x}_{j,1,T-1} \bar{x}_{j,1,T-1} \right)^{-1} \left[ \left( B_j^m \right)^{-1} \psi_j^m + \left( \sigma_j^m \right)^{-2} \bar{x}_{j,1,T-1} \bar{x}_{j,1,T-1} \right], \\
\left[ \left( B_j^m \right)^{-1} + \left( \sigma_j^m \right)^{-2} \bar{x}_{j,1,T-1} \bar{x}_{j,1,T-1} \right]^{-1} \right)
\]

\(^{17}\) Note that \( \alpha_3 \) drawn at Step 1 includes values for \( \Delta c_3, \Delta c_2, \Delta c_1, \) and \( \Delta c_0 \), hence the possibility to compile \( y_j^m(L) \Delta c_3 \) even for \( j = 4 \).
Step 6a: Draw $\sigma_1^q$ conditional on $u_{1,1,T}$ and $\psi_1^q$, based on the following equation:

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

Prior distribution: $(\sigma_1^q)^2 \sim IG(\nu_1^q, f_1^q)$. 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 $f_1^q$.

Posterior distribution:

$$(\sigma_1^q)^2 \sim IG\left(\nu_1^q + \frac{T - 1}{2}, f_1^q + \frac{1}{2} \left(\bar{u}_{1,2,T}^q - \bar{u}_{1,1,T-1}^q \psi_1^q\right)^2 \left(\bar{u}_{1,2,T}^q - \bar{x}_{1,1,T-1}^q \psi_1^q\right)\right)$$

Step 6b: For $j = 1...4$, draw $\sigma_j^m$ conditional on $\Delta y_{j,1,T}^m$, $\Delta c_{1,T}$ and $\psi_j^m$, based on representation 2 the 4 following independent equations:

$$\frac{\Delta y_{j,t}^m - y_{j,t}^m(L) \Delta c_{t}^e}{\equiv x_{j,t}^m} = \frac{\psi_j^m (\Delta y_{j,t-1}^m - y_{j,t-1}^m(L) \Delta c_{t-1}^e)}{\equiv x_{j,t-1}^m \psi_j^m} + \sigma_j^m \cdot \varepsilon_{jt}$$

Prior distributions: $(\sigma_j^m)^2 \sim IG(\nu_j^m, f_j^m)$. 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 $f_j^m$.

Posterior distributions:

$$(\sigma_j^m)^2 \sim IG\left(\nu_j^m + \frac{T - 1}{2}, f_j^m + \frac{1}{2} \left(x_{j,2,T}^m - x_{j,1,T-1}^m \psi_j^m\right)^2 \left(x_{j,2,T}^m - x_{j,1,T-1}^m \psi_j^m\right)\right)$$

Step 7: Draw $\phi_1$ conditional on $\Delta c_{1,T}$, $S_{1,T}$, $V_{1,T}$, and $(\mu_{S1}, V_1, ..., \mu_{St}, V_T)$, based on the following (state) equation:

$${\Delta c_t - \mu_{S_t}, V_t \over \sqrt{1 + h \cdot V_t \cdot \sigma_c}} \equiv z_t \equiv {\Delta c_{t-1} - \mu_{S_{t-1}}, V_{t-1} \over \sqrt{1 + h \cdot V_{t-1} \cdot \sigma_c}} + v_t$$

Prior distribution: $\phi_1 \sim N\left(\phi_{1,}, \Sigma_{1}\right)$. Following Kim and Nelson (1998), we use $\phi_{1,} = 0$ and $\Sigma_{1} = 1$. This prior allows for factor autocorrelation but remains diffuse given its large variance.
Posterior distribution:
\[ \phi_1 | \Delta c_T, S_T, \mu_0, \mu_1 \sim N \left( \left( \frac{1}{\mu_0} + \frac{1}{\mu_1} + \frac{1}{\mu_1} \right)^{-1} \left( \frac{1}{\mu_0} + \frac{1}{\mu_1} \right), \left( \frac{1}{\mu_0} + \frac{1}{\mu_1} \right)^{-1} \right) \]

**Step 8:** Draw \( \mu_{00}, \mu_{01}, \mu_{10} \) and \( \mu_{11} \) conditional on \( \Delta c_{1,T}, S_{1,T}, V_{1,T}, \phi_1, h \) and \( \sigma_C^2 \), based on the following (state) equation:

\[
\frac{\Delta c_t - \phi_1 \Delta c_{t-1}}{\sqrt{1 + h \cdot V_t \cdot \sigma_C}} = \mu_{00} \frac{1}{\sqrt{1 + h \cdot V_t \cdot \sigma_C}} + \mu_{01} \frac{V_t}{\sqrt{1 + h \cdot V_t \cdot \sigma_C}} + \mu_{10} \frac{S_t}{\sqrt{1 + h \cdot V_t \cdot \sigma_C}} + \mu_{11} \frac{S_t \cdot V_t}{\sqrt{1 + h \cdot V_t \cdot \sigma_C}} + \nu_t
\]

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

Following Kim and Nelson (1998), we use \( \tilde{\mu} = 0_{4 \times 1} \) and \( D = I_{d_4} \).

Notations: \( L_T = \frac{\Delta c_T}{\sqrt{1 + h \cdot V_T \cdot \sigma_C}} \); \( R_T = \begin{pmatrix} \frac{1}{\sqrt{1 + h \cdot V_T \cdot \sigma_C}} \\ \frac{V_T}{\sqrt{1 + h \cdot V_T \cdot \sigma_C}} \\ \frac{S_T}{\sqrt{1 + h \cdot V_T \cdot \sigma_C}} \\ \frac{S_T \cdot V_T}{\sqrt{1 + h \cdot V_T \cdot \sigma_C}} \end{pmatrix} \)

**Step 9:** Draw \( h \) conditional on \( \Delta c_{1,T}, S_{1,T}, V_{1,T}, \phi_1, \tilde{\mu} \) and \( \sigma_C^2 \), based on the following (state) equation:

\[
\frac{\Delta c_l}{\sigma_C} = \frac{\mu_{s_c} \cdot V_l}{\sigma_C} + \phi_1 \frac{\Delta c_{l-1}}{\sigma_C} + \sqrt{1 + h \cdot V_l \cdot \nu_l}
\]

Prior distribution: \( 1 + h \cdot IG(v_h, f_h) \). For the shape parameter, we use \( v_h = 5 \), 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:

\[
1 + h \mid \Delta c_{1,T}, \Delta \mu, \phi_1, \mu, \sigma_c \sim IG \left( \frac{v_h + \sum_t 1_{\{V_t=1\}}}{2}, \left( f_h + \frac{\sum_t 1_{\{V_t=1\}} \cdot (\Delta c_t - \mu S_t V_t - \phi_1 \Delta c_{t-1})^2}{2 \sigma_c^2} \right) \right)_{[1+h\geq 1]}
\]

**Step 10:** Draw \( \sigma_c^2 \) conditional on \( \Delta c_{1,T}, \Delta \mu, \phi_1, \mu, \sigma_c \) and \( h \), based on the following (state) equation:

\[
\frac{\Delta c_t}{\sqrt{1 + h \cdot V_t}} = \frac{\mu S_t V_t}{\sqrt{1 + h \cdot V_t}} + \phi_1 \frac{\Delta c_{t-1}}{\sqrt{1 + h \cdot V_t}} + \sigma_c \cdot v_t
\]

Prior distribution: \( \sigma_c^2 \sim IG(v_{\sigma_c}, f_{\sigma_c}) \). For the shape parameter, we use \( v_{\sigma_c} = 2 \), 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 \( f_{\sigma_c} \) 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).

Posterior distribution:

\[
\sigma_c^2 \mid \Delta c_{1,T}, \Delta \mu, V_T, \phi_1, \mu, h \sim IG \left( v_{\sigma_c} + \frac{T}{2}, f_{\sigma_c} + \frac{\sum_t \left( \Delta c_t - \mu S_t V_t - \phi_1 \Delta c_{t-1} \right)^2}{2 \sigma_c^2} \right)
\]

**Step 11:** Draw \( \sigma_a^2 \) conditional on \( \Delta c_{1,T} \) based on the following equation:

\[
a^q_{1,t} = a^q_{1,t-1} + \sigma_a \cdot \eta^a_t
\]

Prior distribution: \( \sigma_a^2 \sim IG(v_{\sigma_a}, f_{\sigma_a}) \). For the shape parameter, we use \( v_{\sigma_a} = 2 \), 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 \( f_{\sigma_a} \) 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 U.S. potential GDP growth rate released by the Congressional Budget Office. 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.

---

18 In the FRED economic database of the Federal Reserve of Saint Louis, the corresponding series has the code GDPPOT.
Posterior distribution:

\[
\sigma^2_a | \gamma_{1:T} \sim IG \left( v_{\sigma_a} + \frac{T}{2} \cdot f_{\sigma_a} + \frac{\sum_t (a^q_{1,t} - a^q_{1,t-1})^2}{2} \right)
\]

**Step 12:** Draw transition probabilities \( PrS_{11} \) and \( PrS_{00} \) conditional on \( S_{1:T} \)

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

Likelihood function: \( L(PrS_{11}, PrS_{00} | S_{1:T}) = PrS_{11}^{n_{11}^S} (1 - PrS_{11})^{n_{10}^S} PrS_{00}^{n_{00}^S} (1 - PrS_{00})^{n_{01}^S} \)

where \( n_{ij}^S \) refer to transitions from state \( i \) to state \( j \) which can be counted based on \( S_{1:T} \).

Posterior distributions:

\( PrS_{11} | S_{1:T} \sim \text{beta}(u^S_{11} + n^S_{11}, u^S_{10} + n^S_{10}) \); \( PrS_{00} | S_{1:T} \sim \text{beta}(u^S_{00} + n^S_{00}, u^S_{01} + n^S_{01}) \)

**Step 13:** Draw transition probabilities \( PrV_{11} \) and \( PrV_{00} \) conditional on \( V_{1:T} \)

Prior distributions: \( PrV_{11} \sim \text{beta}(u^V_{11}, u^V_{10}) \); \( PrV_{00} \sim \text{beta}(u^V_{00}, u^V_{01}) \). We rely on informative priors to guide the identification of volatility states: \( u^V_{00} = 285 \); \( u^V_{01} = u^V_{10} = 1 \); \( u^V_{11} = 285 \). 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: \( L(PrV_{11}, PrV_{00} | V_{1:T}) = PrV_{11}^{n_{11}^V} (1 - PrV_{11})^{n_{10}^V} PrV_{00}^{n_{00}^V} (1 - PrV_{00})^{n_{01}^V} \)

where \( n_{ij}^V \) refer to transitions from state \( i \) to state \( j \) which can be counted based on \( V_{1:T} \).

Posterior distributions:

\( PrV_{11} | V_{1:T} \sim \text{beta}(u^V_{11} + n^V_{11}, u^V_{10} + n^V_{10}) \); \( PrV_{00} | V_{1:T} \sim \text{beta}(u^V_{00} + n^V_{00}, u^V_{01} + n^V_{01}) \)
APPENDIX 5: Smoothed inference on $S_t$ (state of the business cycle: expansion or recession)

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

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

**Result 1:** $p(S_{1-T} | \Delta y_{1-T}^c, \Delta c_{1-T}, V_{1-T}, \theta) = p(S_{1-T} | \Delta c_{1-T}, V_{1-T}, \theta)$

Proof: Once $\Delta c_{1-T}$ is included in the information, $\Delta y_{1-T}^c$ does not bring any additional information on $S_{1-T}$. $p(S_{1-T} | \Delta c_{1-T}, V_{1-T}, \theta)$ can then be rewritten as follows:

$$p(S_{1-T} | \Delta c_{1-T}, V_{1-T}, \theta) = p(S_T | \Delta c_{1-T}, V_{1-T}, \theta) \cdot p(S_{T-1} | S_T, \Delta c_{1-T}, V_{1-T}, \theta) \cdot p(S_{1-T-2} | S_{T-1}, S_T, \Delta c_{1-T}, V_{1-T}, \theta)$$

**Result 2:** $p(S_{T-1} | S_T, \Delta c_{1-T}, V_{1-T}, \theta) = p(S_{T-1} | S_T, \Delta c_{1-T-1}, V_{1-T}, \theta)$

Proof:

$$p(S_{T-1} | S_T, \Delta c_{1-T-1}, V_{1-T}, \theta) = \frac{p(\Delta c_T | S_{T-1}, S_T, \Delta c_{1-T-1}, V_{1-T}, \theta) \cdot p(S_{T-1} | S_T, \Delta c_{1-T-1}, V_{1-T}, \theta)}{p(\Delta c_T | S_T, \Delta c_{1-T-1}, V_{1-T}, \theta)} = p(S_{T-1} | S_T, \Delta c_{1-T-1}, V_{1-T}, \theta)$$

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

$$p(\Delta c_T | S_{T-1}, S_T, \Delta c_{1-T}, V_{1-T}, \theta) = p(\Delta c_T | S_T, \Delta c_{1-T}, V_{1-T}, \theta)$$

**Result 3:** $p(S_{1-T-2} | S_{T-1}, S_T, \Delta c_{1-T}, V_{1-T}, \theta) = p(S_{1-T-2} | S_{T-1}, S_T, \Delta c_{1-T-2}, V_{1-T}, \theta)$

Proof:

$$p(S_{1-T-2} | S_{T-1}, S_T, \Delta c_{1-T}, V_{1-T}, \theta) = \frac{p(S_{1-T-2}, \Delta c_{T-1}, \Delta c_T | S_{T-1}, S_T, \Delta c_{1-T-2}, V_{1-T}, \theta)}{p(\Delta c_{T-1}, \Delta c_T | S_{T-1}, S_T, \Delta c_{1-T-2}, V_{1-T}, \theta)} = \frac{p(\Delta c_{T-1}, \Delta c_T | S_{1-T-2}, S_{T-1}, S_T, \Delta c_{1-T-2}, V_{1-T}, \theta) \cdot p(S_{1-T-2} | S_{T-1}, S_T, \Delta c_{1-T-2}, V_{1-T}, \theta)}{p(\Delta c_{T-1}, \Delta c_T | S_{T-1}, S_T, \Delta c_{1-T-2}, V_{1-T}, \theta)} = p(S_{1-T-2} | S_{T-1}, S_T, \Delta c_{1-T-2}, V_{1-T}, \theta)$$

45
The last equality follows from the fact that $\Delta c_{T-1}$ and $\Delta c_T$ do not depend on $S_{t-2}$, hence:

$$p(\Delta c_{T-1}, \Delta c_T|S_{t-2}, S_{T-1}, S_T, \Delta c_{1:T-2}, V_{1:T}, \theta) = p(\Delta c_{T-1}, \Delta c_T|S_{T-1}, S_T, \Delta c_{1:T-2}, V_{1:T}, \theta)$$

**Result 4:** $p(S_{1:T-2}|S_{T-1}, S_T, \Delta c_{1:T-2}, V_{1:T}, \theta) = p(S_{1:T-2}|S_{T-1}, \Delta c_{1:T-2}, V_{1:T}, \theta)$

Proof:

$$p(S_{1:T-2}|S_{T-1}, S_T, \Delta c_{1:T-2}, V_{1:T}, \theta)$$

$$= p(S_T|S_{1:T-2}, S_{T-1}, \Delta c_{1:T-2}, V_{1:T}, \theta) \cdot p(S_{1:T-2}|S_{T-1}, \Delta c_{1:T-2}, V_{1:T}, \theta)$$

$$= p(S_{1:T-2}|S_{T-1}, \Delta c_{1:T-2}, V_{1:T}, \theta)$$

The last equality follows from the fact that

$$p(S_T|S_{1:T-2}, S_{T-1}, \Delta c_{1:T-2}, V_{1:T}, \theta) = p(S_T|S_{T-1}, \Delta c_{1:T-2}, V_{1:T}, \theta)$$

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

$$p(S_{1:T}|\Delta y_{1:T}, \Delta c_{1:T}, V_{1:T}, \theta)$$

$$= p(S_T|\Delta c_{1:T}, V_{1:T}, \theta) \cdot p(S_{T-1}|S_T, \Delta c_{1:T-1}, V_{1:T}, \theta) \cdot p(S_{1:T-2}|S_{T-1}, \Delta c_{1:T-2}, V_{1:T}, \theta)$$

$$= p(S_T|\Delta c_{1:T}, V_{1:T}, \theta) \cdot \prod_{t=1}^{T-1} p(S_t|S_{t+1}, \Delta c_{1:T}, V_{1:T}, \theta)$$

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

$$p(S_t = j|S_{t+1} = k, \Delta c_{1:T}, V_{1:T}, \theta)$$

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

$$= \frac{\sum_j p(S_{t+1} = k|S_t = j, \Delta c_{1:T}, V_{1:T}, \theta) \cdot p(S_t = j|\Delta c_{1:T}, V_{1:T}, \theta)}{\sum_j p(S_{t+1} = k|S_t = j) \cdot p(S_t = j|\Delta c_{1:T}, V_{1:T}, \theta)}$$
APPENDIX 6: Marginal likelihood computation of Markov-Switching Dynamic Factor Models

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:

\[
\log f(\Delta y_{1,T} | \tilde{\theta}) = \log f(\Delta y_{1,T} | \tilde{\theta}^*) + \log \varphi(\tilde{\theta}^*) - \log \varphi(\tilde{\theta}^* | \Delta y_{1,T})
\]  

(2)

A5.1. Evaluating \( f(\Delta y_{1,T} | \tilde{\theta}^*) \)

In the case of a MS-DFM, \( f(\Delta y_{1,T} | \tilde{\theta}^*) \) cannot be evaluated analytically, but the Kalman filter step of the Gibbs Sampling algorithm provides draws from another conditional log-likelihood, namely \( \log f(\Delta y_{1,T} | S_{1,T}^*, V_{1,T}^*, \tilde{\theta}^*) \), where the conditioning set includes the two Markov-Switching variables \( S_{1,T}^* \) and \( V_{1,T}^* \), 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:

\[
\log f(\Delta y_{1,T} | \tilde{\theta}^*) = \sum_{t=1}^{T} \log f(\Delta y_{t} | \Delta y_{1,t-1}^*, \tilde{\theta}^*)
\]

(3)

\[
\log f(\Delta y_{1,T} | S_{1,T}^*, V_{1,T}^*, \tilde{\theta}^*) = \sum_{t=1}^{T} \log f(\Delta y_{t} | \Delta y_{1,t-1}^*, S_{1,T}^*, V_{1,T}^*, \tilde{\theta}^*) = \sum_{t=1}^{T} \log f(\Delta y_{t} | \Delta y_{1,t-1}^*, S_{1,T}^*, V_{1,T}^*, \tilde{\theta}^*)
\]

(4)

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

\[
f(\Delta y_{t} | \Delta y_{1,t-1}^*, \tilde{\theta}^*) = \int f(\Delta y_{t} | \Delta y_{1,t-1}^*, S_{1,T}^*, V_{1,T}^*, \tilde{\theta}^*) \cdot g(S_{1,T}^*, V_{1,T}^*, \Delta y_{1,t-1}^*, \tilde{\theta}^*) \cdot dS_{1,t} \cdot dV_{1,t}
\]

(5)

At first sight, it might seem that this integral cannot be evaluated because \( (S_{1,T}^*, V_{1,T}^*, \Delta y_{1,t-1}^*) \) is not known analytically, contrary to \( f(\Delta y_{t} | \Delta y_{1,t-1}^*, S_{1,T}^*, V_{1,T}^*, \tilde{\theta}^*) \) which is an output of the Kalman filter.

\[\text{Note that } \tilde{\theta}^* \text{ includes all constant model parameters, but no Markov-Switching variable. It is indeed well known that Chib's (1995) formula gives more accurate estimates of model marginal likelihoods if the number of variables in the information set is limited to the maximum extent.}\]
Nevertheless, the integral can be numerically approximated by drawing from $g(S_{1,t}, V_{1,t}, \Delta y_{1,t-1}, \tilde{t}^*)$ using a particle filter.

**A5.1a. Particle filter**

The particle filter applies the logic of sequential importance sampling\(^{20}\). Instead of drawing from $g(S_{1,t}, V_{1,t}, \Delta y_{1,t}, \tilde{t}^*)$ 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, $g(S_{1,t}, V_{1,t}, \Delta y_{1,t}, \tilde{t}^*)$ is approximated with an importance distribution $q_{1,t}(S_{1,t}, V_{1,t}, \Delta y_{1,t}, \psi)$ taking a recursive form:

$$q_{1,t}(S_{1,t}, V_{1,t}, \Delta y_{1,t}, \psi) \equiv q_t(S_t, V_t|S_{1,t-1}, V_{1,t-1}, \Delta y_{1,t-1}, \psi) \cdot q_{1,t-1}(S_{1,t-1}, V_{1,t-1}|\Delta y_{1,t-1}, \psi) \quad (6)$$

**A5.1b. Particle weights**

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

$$f(\Delta y_t|\Delta y_{1,t-1}, \tilde{t}^*) = \int f(\Delta y_t|\Delta y_{1,t-1}, S_{1,t}, V_{1,t}, \tilde{t}^*) \cdot g(S_{1,t}, V_{1,t}, \Delta y_{1,t}, \tilde{t}^*) \cdot dS_{1,t}^{\gamma} \cdot dV_{1,t}^{\gamma}$$

$$= \int f(\Delta y_t|\Delta y_{1,t-1}, S_{1,t}, V_{1,t}, \tilde{t}^*) \cdot p(S_t, V_t|S_{1,t-1}, V_{1,t-1}, \tilde{t}^*) \cdot \frac{g(S_{1,t-1}, V_{1,t-1}, \Delta y_{1,t-1}, \tilde{t}^*)}{q_{1,t-1}(S_{1,t-1}, V_{1,t-1}|\Delta y_{1,t-1}, \psi)}$$

$$\cdot q_{1,t-1}(S_{1,t-1}, V_{1,t-1}|\Delta y_{1,t-1}, \psi) \cdot dS_{1,t}^{\gamma} \cdot dV_{1,t}^{\gamma} \quad (7)$$

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

$$W_t(S_{1,t}, V_{1,t}) = \frac{h(S_{1,t}, V_{1,t}, \Delta y_{1,t}, \tilde{t}^*)}{\int h(S_{1,t}, V_{1,t}, \Delta y_{1,t}, \tilde{t}^*) \cdot dS_{1,t}^{\gamma} \cdot dV_{1,t}^{\gamma}} \cdot \frac{1}{Z_t}$$

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

\(^{20}\) 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.
\[ Z_t \equiv \int h(S_{1:t}^- V_{1:t}^- \Delta y_{1:t}^\prime \hat{\theta}^*) \cdot dS_{1:t}^- \cdot dV_{1:t}^- = \int \frac{h(S_{1:t}^- V_{1:t}^- \Delta y_{1:t}^\prime \hat{\theta}^*)}{q_{1:t}(S_{1:t}^- V_{1:t}^- | \Delta y_{1:t}^\prime, \hat{\psi})} \cdot q_{1:t}(S_{1:t}^- V_{1:t}^- | \Delta y_{1:t}^\prime, \hat{\psi}) \cdot dS_{1:t}^- \cdot dV_{1:t}^- \] (9)

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 \( q_{1:t}(S_{1:t}^- V_{1:t}^- | \Delta y_{1:t}^\prime, \hat{\theta}^*) \). In the following, we will thus focus on unnormalised weights.

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

\[
w_t(S_{1:t}^- V_{1:t}^-) = \frac{h(S_{1:t-1}^- V_{1:t-1}^- \Delta y_{1:t-1}^\prime \hat{\theta}^*)}{h(S_{1:t-1}^- V_{1:t-1}^- \Delta y_{1:t-1}^\prime \hat{\theta}^*)} \cdot q_{1:t}(S_{1:t}^- V_{1:t}^- | S_{1:t-1}^- V_{1:t-1}^-, \Delta y_{1:t}^- \hat{\psi}) \cdot \frac{h(S_{1:t-1}^- V_{1:t-1}^- \Delta y_{1:t-1}^\prime \hat{\theta}^*)}{q_{1:t-1}(S_{1:t-1}^- V_{1:t-1}^- | \Delta y_{1:t-1}^\prime, \hat{\psi})} \equiv \alpha_t(S_{1:t}^- V_{1:t}^-) \] (10)

The transition factor \( \alpha_t(S_{1:t}^- V_{1:t}^-) \) between unnormalised weights at dates (t-1) and t can be rewritten as follows:

\[
a_t(S_{1:t}^- V_{1:t}^-) = \frac{p(S_t, V_t, \Delta y_t | S_{1:t-1}^- V_{1:t-1}^\prime, \Delta y_{1:t-1}^\prime, \hat{\theta}^*)}{q_t(S_{1:t}^- V_{1:t}^- | S_{1:t-1}^- V_{1:t-1}^\prime, \Delta y_{1:t}^\prime, \hat{\psi})} = \frac{f(\Delta y_t | \Delta y_{1:t-1}^\prime, S_{1:t}^- V_{1:t}^-, \hat{\theta}^*) \cdot p(S_t, V_t | S_{1:t-1}^- V_{1:t-1}^\prime, \Delta y_{1:t-1}^\prime, \hat{\theta}^*)}{q_t(S_{1:t}^- V_{1:t}^- | S_{1:t-1}^- V_{1:t-1}^\prime, \Delta y_{1:t}^\prime, \hat{\psi})} \Rightarrow a_t(S_{1:t}^- V_{1:t}^-) = \frac{f(\Delta y_t | \Delta y_{1:t-1}^\prime, S_{1:t}^- V_{1:t}^-, \hat{\theta}^*) \cdot p(S_t, V_t | S_{1:t-1}^- V_{1:t-1}^\prime, \Delta y_{1:t-1}^\prime, \hat{\theta}^*)}{q_t(S_{1:t}^- V_{1:t}^- | S_{1:t-1}^- V_{1:t-1}^\prime, \Delta y_{1:t}^\prime, \hat{\psi})} \] (11)

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

**A5.1c. Resampling**

In the present setting where the time horizon is long (close to 600 months), it is very likely that some particles get a disproportionated 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:

\[
ESS_t = \left( \sum_{i=1}^{N} W_{it}^2 \right)^{-1} \] (12)
It takes values between 1 (when a single particle gets all the weight) and N (when particles are equally weighted).

**A5.1d. 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:

$$q_t(S_t, V_t | S_{1:t-1}, V_{1:t-1}, \Delta y_{1:t}, \hat{y}) \equiv p(S_t, V_t | S_{1:t-1}, V_{1:t-1}, \Delta y_{1:t}, \hat{y}^*)$$  \hspace{1cm} (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 \(f(\Delta y_t | \Delta y_{1:t-1}, S_{1:t-1}, V_{1:t-1}, \hat{y}^*)\) returned by the Kalman filter:

$$p(S_t, V_t | S_{1:t-1}, V_{1:t-1}, \Delta y_{1:t}, \hat{y}^*) = \frac{f(\Delta y_t | \Delta y_{1:t-1}, S_{1:t-1}, V_{1:t-1}, \hat{y}^*) \cdot p(S_t, V_t | S_{1:t-1}, V_{1:t-1}, \Delta y_{1:t-1}, \hat{y}^*)}{f(\Delta y_t | \Delta y_{1:t-1}, S_{1:t-1}, V_{1:t-1}, \hat{y}^*)}$$

$$= \frac{f(\Delta y_t | \Delta y_{1:t-1}, S_{1:t-1}, V_{1:t-1}, \hat{y}^*) \cdot p(S_t = i, V_t = j | S_{1:t-1}, V_{1:t-1}, \Delta y_{1:t-1}, \hat{y}^*)}{f(\Delta y_t | \Delta y_{1:t-1}, S_{1:t-1}, V_{1:t-1}, \hat{y}^*)}$$ \hspace{1cm} (14)

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

$$\alpha_t(S_{1:t}, V_{1:t}) = f(\Delta y_t | \Delta y_{1:t-1}, S_{1:t-1}, V_{1:t-1}, \hat{y}^*)$$

$$= \sum_{i,j} f(\Delta y_t | \Delta y_{1:t-1}, S_{1:t-1}, V_{1:t-1}, S_t = i, V_t = j, \hat{y}^*) \cdot p(S_t = i, V_t = j | S_{1:t-1}, V_{1:t-1}, \Delta y_{1:t-1}, \hat{y}^*)$$ \hspace{1cm} (15)

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

**A5.2. Evaluating \(\varphi(\hat{y})\) and \(\varphi(\hat{y}^*|\Delta y_{1:t})\)**

Here we follow the strategy advocated by Bauwens and Rombouts (2012). The key to evaluate \(\varphi(\hat{y}^*)\) and \(\varphi(\hat{y}^*|\Delta y_{1:t})\) is to split the vector \(\hat{y}^*\) into the same subvectors \((\hat{\theta}_1, ..., \hat{\theta}_\kappa)\) as in the Gibbs sampling algorithm. In this case, the prior density \(\varphi(\hat{y}^*)\) can be evaluated using \(K\) analytically known distributions, as follows:

$$\varphi(\hat{y}^*) = \prod_{i=1}^{\kappa} \varphi(\hat{\theta}_i^*)$$ \hspace{1cm} (16)
The posterior density $\varphi(\hat{\theta}^* | \hat{\Delta}y_{1,T})$ can also be factorised into $K$ terms corresponding to the different Gibbs sampling steps:

$$ \varphi(\hat{\theta}^* | \hat{\Delta}y_{1,T}) = \varphi(\hat{\theta}_1^* | \hat{\Delta}y_{1,T}) \cdot \varphi(\hat{\theta}_2^* | \hat{\theta}_1^*, \hat{\Delta}y_{1,T}) \cdot \ldots \cdot \varphi(\hat{\theta}_K^* | \hat{\theta}_1^*, \ldots, \hat{\theta}_{K-1}^*, \hat{\Delta}y_{1,T}) $$  \hspace{1cm} (17)

Each of these terms is finally evaluated using $G$ draws from an auxiliary Gibbs sampler. For example, $\varphi(\hat{\theta}_2^* | \hat{\theta}_1^*, \hat{\Delta}y_{1,T})$ is evaluated as:

$$ \frac{1}{G} \sum_{i=1}^{G} \varphi(\hat{\theta}_2^* | \hat{\theta}_1^*, \hat{\theta}_3^*, \ldots, \hat{\theta}_K^*, S_{1,T}^i, V_{1,T}^i, \hat{\Delta}y_{1,T}) $$

where $\hat{\theta}_1^*$ is the posterior mean of the first subvector of parameters, and $(\hat{\theta}_3^*, \ldots, \hat{\theta}_K^*, S_{1,T}^i, V_{1,T}^i)$ are draws from an auxiliary Gibbs sampler which is similar to the one used in the main text, except that $\hat{\theta}_1^*$ is set equal to $\hat{\theta}_1^*$. The same reasoning applies for the other terms in equation (17).