# Changing correlations, macroeconomic risk and the business cycle

Lorena Saiz*

European Central Bank

The recession that started by the end of 2007 in the United States and later on in Europe, popularly known as the *great recession*, has put a question mark on some traditional methods for forecasting, in particular dynamic factor models. These models exploit the comovement of a multiple set of variables to predict macroeconomic aggregates such as inflation or GDP growth. But this episode has demonstrated that the comovement story is incomplete when the dynamics of the variances are not considered.

We show that introducing time-varying variances in traditional Dynamic Factor Models helps not only to solve estimation problems but can also improve forecasts. Furthermore, it has interesting implications for the conditional correlations that are also time varying. In addition, we get an estimate of the common factor's volatility which in the business cycles context can be interpreted as a measure of broad macroeconomic risk.

**Keywords**: Stochastic volatility, Sequential Monte Carlo simulation, Rao-Blackwellised particle filter, Kalman filter, Business cycles.

**JEL classification**: C150, C530, C580.

## 1. Introduction

The recession that started by the end of 2007 in the United States and later on in Europe, popularly known as the *great recession*, has put a question mark on some traditional

*E-mail address: Lorena.Saiz@ecb.int. I am grateful to Gabriel Pérez-Quirós, Eva Senra, Dolores Gadea, Pilar Poncela, Juan F. Jimeno, Jesús Crespo-Cuaresma and Tomás Mancha for valuable comments and suggestions. The views expressed in this paper are those of the author and do not necessarily reflect those of the ECB or the Eurosystem.

methods for forecasting. The majority of these models were incapable of predicting the timing and also the magnitude of the recession. With the high uncertainty about the recovery and indicators that were sometimes misleading, forecasting has become more challenging. As a matter of fact, the magnitude of the recession was so important that it is comparable to the recessions of the 1970s and so unusually big for the recent times that it has become very complicated to make predictions about the future with traditional methods. Some people have included financial variables in their macro models to account for the volatility, because the recession started as a financial crisis and the financial sector has a more important role than ever. However, there is increasingly a consensus that what is crucial is to take into account the changes in the level of uncertainty or volatility in models to predict macro variables such as GDP or inflation. In other words, macroeconomics and volatility are closely related because finance and macroeconomics are more interconnected than ever.

After the double-dip recession in the USA in the beginning of the 1980s, there was a process of declining variance of the main macroeconomic aggregates called the *great moderation*. Since then, the cycles were smoother, especially the recessions, but also the recoveries were milder. The same smoothing phenomenon was observed in Europe, although not as markedly. However, the last recession involved a sharp decline in output growth. In the specific case of the Euro area, where the available historical data set is not as long as for the US, the small sample size aggravates the problem because the impact of the most recent observations is inversely proportional to the length of the data that precede it. All the models, and in particular the Dynamic Factor models (DFM), failed to predict the great recession. The key question is why they were not able to do it.

There is a consensus in the economic literature about some stylized facts for key macroeconomic time series. First, there is a strong *comovement* or *correlation* among most of the macroeconomic indicators, especially at business cycles frequencies (from the earliest works by Burns and Mitchell (1947), Geweke (1977) and Sargent and Sims (1977), to the most recent ones which exploit this comovement by means of factor models to predict GDP growth, among others Camacho and Pérez-Quirós (2010) or Angelini et al. (2011)). This explains the popularity of factor models to estimate and forecast the business cycle. Second, in crises episodes the volatility of most indicators increases (see Engle (1982) and Stock and Watson (2007) for inflation; Weiss (1984) and Ewing and Thompson (2008) for industrial production; Ho and Tsui (2003) and Fang, Miller and Lee (2008) for GDP growth). This is what is called *volatility clustering* or non-constancy of the conditional variance over time. Third, the aforementioned phenomenon occurs simultaneously in many macroeconomic series. Therefore, we can conclude that

there exists *volatility comovement* (related works: Bollerslev (1990) for exchange rates; Cappielo, Engle and Sheppard (2003) for equities and bonds; Ho, Tsui and Zhang (2009) for sectoral industrial production). And finally, given that the volatility is higher during recessions than expansions, there is a *leverage* or *asymmetric effect* because negative shocks have a higher impact on volatility than positive shocks (see for instance Ho and Tsui (2003) and Fang , Miller and Lee (2008) for GDP growth, Ewing and Thompson (2008) for industrial production, and Ho, Tsui and Zhang (2009) for sectoral industrial production). Many models, such as the popular DFMs, have focused on the first stylized fact without considering the others. But the recent crisis has shown how important it is to take them into account.

In his recent book "Anticipated correlations" (2009) Robert Engle indeed pointed out that a model which does not update volatilities and correlations will make much bigger mistakes when the markets are changing. The literature on business cycles has focused on the comovement and the non-linear behavior of the conditional mean in order to identify recessions and expansions, without explicitly considering time-varying features of higher order moments (e.g. conditional variances or correlations). However, conditional heteroscedasticity, asymmetric volatility and time-varying conditional correlations have important implications for business cycle theory and especially for forecasting. Furthermore, Nelson and Foster (1994) argued that phenomena like fat residuals or leverage effects are potentially more important than misspecifying conditional means. For all these reasons we think that the comovement story is incomplete when the dynamics of the second order moments are not considered.

The aim of this paper is to study in which ways we can modify the standard DFMs to take into account the non-constancy of second conditional moments (i.e. the variance and the correlations). Our point is to take this instability into account in the most parsimonious way by introducing time-varying variances in the common factor of the DFM. As we will discuss later, this involves that the conditional correlations are also time varying. We will show that DFMs under serial heteroscedasticity do a better job at forecasting than under homoscedasticity. Our empirical strategy entails several advantages: i) we get an estimate of the volatility as a by-product; ii) we take into account the changing level of uncertainty in the confidence intervals for the conditional mean forecasts; and iii) the implicit conditional correlations are time-varying. However, this comes at the cost of introducing an additional source of error, because in some cases the parameters of the volatility models are difficult to estimate.

The structure of the paper is as follows. Section 2 reviews the literature on how works in macroeconomics have dealt with the volatility problem. In Section 3, the DFMs and

possible extensions to account for time-varying variances are presented. The results of a Monte Carlo experiment are explained in Section 4, and Section 5 presents Stock and Watson's (1991) coincident indicator as empirical application. Section 6 concludes.

# 2. Macroeconomics and volatility

## 2.1. Some stylized facts

The data we are going to use in this work correspond to the four monthly indicators that Stock and Watson (1991) used to construct a coincident indicator for the United States and that the National Bureau of Economic Research (NBER) closely monitors to date peaks and troughs in the economic activity. These indicators are: i) The industrial production index which is a monthly measure of production. The main shortcoming is that it only represents manufacturing, mining and utilities sectors, excluding services and government sectors. ii) The real personal income less transfers, which is the monthly measure closest to real Gross Domestic Product (GDP) precisely due to the fact that the transfers are subtracted and the nominal series is deflated with the interpolated quarterly GDP implicit price deflator. iii) The employment series which is typically monitored and included in Stock and Watson's indicator is the number of non-farm payroll employees. It consists of the number of filled jobs in the business sector excluding agriculture and is based on the Current Employment Survey (CES). iv) Finally, the real manufacturing and trade sales is another monthly indicator of output, although it only includes sales of goods and imported goods but not services. In order to be comparable to real GDP, the nominal sales are deflated in the same way than the personal income.

Figure 1 represents the rates of growth and the estimated volatilities using simple GARCH(1,1) models for the quarterly real GDP and the four monthly variables of Stock and Watson's (1991) coincident indicator for the US mentioned above: industrial production, real personal income less transfer payments, real manufacturing and trade sales, and non-farm payroll employment. We see that the majority of these indicators exhibit a higher conditional variance during recessions, especially in the recessions of the 1970s and the last one in 2007. We also observe a global reduction in total variance since the mid-1980s (i.e. the great moderation).

In the top of Figure 2 we plot the absolute value of the determinant of the correlation matrix as a measure of linear relationship for these series. It will be 0 if the variables are perfectly and positively correlated, and 1 whenever the variables are not correlated at all. In order to make this measure more intuitive, we represent it as 1 minus this determinant. In this way, values close to 1 indicate highly correlated series and 0 the

opposite. It is computed using a rolling window of 5 years over the sample from 1959 to 2011. We observe that the total mass of correlation in general has decreased over time as it had passed from values around 0.8 to around 0.3. There is a turning point around the mid-1980s, at the time of the great moderation, but during recessions (shaded areas are the officially dated recessions by the NBER) the correlations increase, especially during the great recession.

In the bottom of Figure 2 we show the same measure of global comovement as before but this time for the estimated volatilities. The graph shows that there exists a certain comovement between the volatilities, in particular during recessions. What we can conclude from all this is that the instability and asymmetries in variance are transmitted to the correlations in a way that times of common high volatility coincide with times of high correlation. Therefore, the correlations are also time-varying.

## 2.2. The importance of variance

As Hamilton (2008) stated, most macroeconomic studies have concentrated their efforts on the conditional mean disregarding the information contained in the conditional variance. However, there is an increasing number of more recent works that stress the importance of accounting for variance instability for the reliability of the results.

The most well-known problem caused by non-constancy of the variance is the *inefficiency* of the estimates. The estimation of linear models by Maximum Likelihood (ML henceforth) is consistent even when GARCH effects in the disturbance are ignored (Weiss, 1984), but the estimators will not be efficient any more. Hamilton (2008) shows that, even when our main interest is the conditional mean, modeling the conditional variance correctly is important for two reasons: i) OLS standard errors can be misleading and result in spurious regressions because we strongly reject a true null hypothesis (i.e. type I error); ii) the inference about the conditional mean can be badly influenced by outliers and high-variance episodes.

Another issue are *identification problems*. Fiorentini and Sentana (2001) found that if factor models are estimated without taking time variation in the conditional variances into account, neither the loading matrix nor the noise variance are identified without extra restrictions. However, the identification problems are somehow alleviated when this variation is accounted for, as there is a relative efficiency gain which increases with the variability of conditional variances.

In estimation methods like ML there could be *convergence problems* due to a complicated shape of the likelihood with multiple modes and some flatter regions. This could even result in Heywood cases, a case in which ML accommodates the parameters to reach

maximum in-sample fit, with absurd parameter values such as negative or zero variances. The likelihood can also have a point mass at zero values of the parameters (Stock and Watson, 1998), which is known as pile-up problem. Usually this occurs when there are convergence problems because the algorithm is going through points in a region of non-identification. On the other hand, as pointed out by Box-Steffensmeier and Lebo (2008 ), in methods like rolling window estimation or the popular Kalman filter, the most volatile period dominates the estimations making the model lose memory. In the rolling window case it is because the rolling estimation introduces a small sample bias. In the Kalman filter estimation, as we will see later, it is due to the construction of the filter.

## 2.3. Literature review

There is an agreement in the literature that mean shifts are less important for the variance than variance shifts are for the mean. In this respect, some works in the literature of structural break tests in mean have noticed that the distribution of the test statistic may be affected. Two of the proposed solutions are to consider either bootstrap p-values adapted to the instability of variance for these tests (e.g. Blocks or Wild bootstrap) or Bayesian mixture innovation models (Gerlach et al. (2000) and Giordani and Kohn (2008)).

Regarding the literature based on unobserved component (UC) models, Bos and Koopman (2010) notice the relevance of modeling the mean and the variance at the same time in an univariate UC model for industrial production. They introduce stochastic volatility to improve the fit of the model because in this way observations from non-volatile periods receive more weight.

In the VAR and SVAR literature the most influential works which have introduced time-varying variance in this framework are those trying to explain the reasons for the great moderation in the US. Basically there are two theories: good policy vs good luck. The good policy approach argues that changes in the transmission mechanism of monetary policy played a role in the great moderation. In contrast, the good luck story focuses on the reduction of the volatility of the shocks. From a practical point of view, for disentangling both effects it is necessary to extend the VAR/SVAR models in a way that the variance of the shocks is time-varying and the impulse responses or covariances also change over time, as in Primiceri (2005) and Koop et al. (2009).

In relation to the identification of turning points, Chauvet and Potter (2010) demonstrated with Bayesian probits that the best model to predict turning points in the economic activity for the US is one with recurrent breaks in variance, and Creal, Koopman and Zivot (2010) show that the beginning of the recession in 2007 is better identified

when they introduce stochastic volatility in their multivariate trend-cycle decomposition model.

For forecasting, most of the works that combine mean and variance modeling at the same time are at an univariate level. Espasa et al. (2010) at a univariate level and Alessi et al. (2009) at a multivariate level show that identifying the source of uncertainty helps.

When we turn to DFMs, there are very few works that have extended these models to account for changes in variance. There are different ways in which the variance could adapt over time. On the one hand, it could be observation-driven, such as in GARCH models, where the conditional variance depends on past observations. Alternatively, it could be parameter-driven, such as in Stochastic Volatility (SV) models, because the conditional variance depends on a latent (or unobserved) component estimated in a state space model. The latter is more flexible although more complicated to estimate. It is important to notice that these time-varying parameter (TVP) models work well when the changes in the parameters are gradual ("many small breaks"). Large and infrequent breaks, such as the great moderation, are more difficult to pick up with these models, especially with the observation-driven approach.

There are two influential papers based on the observation-driven approach. First, Harvey, Ruiz and Sentana's (1992) static factor model including ARCH for the variance of common and idiosyncratic factors. They show the importance of including a correction term in the estimation that accounts for the estimation error in the unobserved factor. Second, Alessi, Barigozzi and Capasso (2009) estimated a dynamic factor model distinguishing between static and dynamic factors. The static factors are linear combinations of dynamic factors and are orthogonal. In contrast, the dynamic factors (or common shocks) are correlated. This fact allows to exploit the multivariate dimension to account for time-varying variances in the model as the variance-covariance matrix of the dynamic factors will follow a multivariate GARCH model[1]. Therefore, the static factors will be weak GARCH processes. The variance of the idiosyncratic factors will follow independent and univariate GARCH models. In the empirical example, they demonstrated that this approach leads to a small improvement in forecasting inflation at different horizons. As a by-product they get estimates and forecast of the conditional volatility and the covariances. Their model can be considered a special case of the structural ARCH developed by Harvey et al. (1992). But notice that this paper belongs to a growing literature on multivariate volatility models which are related to the conditional heteroscedastic DFM. The philosophy of multivariate volatility models[2] is slightly different from the traditional

[1] They consider two possibilities: a BEKK model and a Dynamic Conditional Correlation (DCC) model.
[2] See a survey on multivariate GARCH models in Bauwens, Laurent and Rombouts (2006).

DFM as their purpose is to estimate and explain conditional variance-covariance matrices with the observed variables. But Sentana (1998) demonstrated that under certain conditions, they are observationally equivalent.

Regarding the parameter-driven approach, a recent paper by del Negro and Otrok (2008) estimates from a fully Bayesian perspective a DFM in which all parameters are time varying. The variances follow geometric random walks and the loading coefficients are random walks. In this framework many strong assumptions are required in order to identify the model. In contrast, Stock and Watson (2010) proposed a DFM with stochastic volatility in the variance of common and idiosyncratic factors. They account for the large break in variance of the great moderation by introducing dummy variables. In the next section we will provide more insights about how to extend traditional DFM to account for variance instability.

# 3. Dynamic factor models and volatility

## 3.1. Dynamic factor models

Traditional factor models (e.g. the static factor model, the generalized dynamic factor model, etc.) have been successfully employed to forecast the conditional mean of macroeconomic variables such as inflation or GDP growth. Typically, they assume that there are some common unobserved factors that help to explain the comovement of the series, although this does not explain the whole behavior of the series. There is a part that is specific to each series which is the idiosyncratic component. In order to identify these components some assumptions are required. Here we present the state-space representation of a simple factor model:

$$y_t = Hf_t + \xi_t, \tag{3.1}$$

$$f_t = Ff_{t-1} + u_t \tag{3.2}$$

with $\xi_t \backsim N(0, R)$ and $u_t \backsim N(0, Q)$, where $y_t = \{y_{1t}, ..., y_{Nt}\}$ is the vector of $N$ series for $t = 1, ..., T$; $f_t = \{f_{1t}, .., f_{kt}\}$ is the vector of $K$ common factors that follow an autoregressive model (for simplification we assume $p = 1$); and $\xi_t = \{\xi_{1t}, ..., \xi_{Nt}\}$ is the vector of $N$ idiosyncratic components. To simplify we do not introduce dynamics in this component. $H$ is a $NxK$ matrix with the factor loadings in the measurement 3.1, $F$ is a $KxK$ matrix of the autoregressive coefficients for the transition 3.2, and $R$ and $Q$ are the variance-covariance matrices of the idiosyncratic components and the innovations in the

factor equation. Both are diagonal matrices of dimensions $NxN$ and $KxK$, respectively.

To identify the unobserved components the following assumptions are usually made: i) $u_t$ is an orthonormal white noise involving that $var(u_t) = Q = I$. This normalization assumption is crucial to identify the loadings and factors; ii) $\xi_{it}$ is a zero-mean stationary process and independent across i; iii) $\xi_{it-k}$ and $u_{jt}$ are mutually orthogonal (independent), for all integer $k$, $i$, and $j$ (i.e. exact factor model). This orthogonality assumption is sometimes relaxed to allow for a limited amount of cross-sectional correlation (i.e. approximate factor model). Assumptions ii) and iii) guarantee that all the comovement comes from the common factors and make it possible to identify common and idiosyncratic factors. Notice that this model implies the variance-covariance decomposition

$$var(y_t) = \Sigma = Hvar(f_t)H' + R \tag{3.3}$$

Regarding the estimation method, Maximum Likelihood with Kalman filter[3] is typically the preferred method for small or moderate $N$, and Static and Dynamic Principal Components Analysis (PCA and DPCA) for large $N$. Fiorentini and Sentana (2001) showed that by estimating these models without taking time variation in the conditional variances into account, neither the loading matrix nor the noise variance were identified without extra restrictions. However, if only some factors have conditional heteroscedastic variances and we take it into account in the estimation, the loading matrix is identifiable. Therefore, the identification problems are alleviated when variation in factor variances is accounted for because there is a relative efficiency gain to estimate the loading matrix and the variances of the idiosyncratic components, which increases with the variability of the conditional variances. Besides, the indeterminacy of factors is small when the variance of idiosyncratic components is small compared to the variance of the common component, $Hvar(f_t)H'$. This means that in the hypothetical case that $Q = I$ and $F = I$, if the trace of $HR^{-1}H'$ is large, the indeterminacy of the factor is small. Notice that in this case the inverse of the idiosyncratic variance-covariance matrix $R^{-1}$ is also the inverse of the signal-to-noise ratio, and $H$ will be unique if $HR^{-1}H'$ is diagonal.

An efficient method to estimate the state variables given the parameters is the Kalman filter. This filter estimates the state variables as a weighted average of the most recent observed values of the variables together with the past values, $f_{t+1|t} = \sum_{j=1}^{t} w_j(f_{t+1|t})y_j$. Koopman and Harvey (2003) derived the weights $w_j(f_{t+1|t})$ of the Kalman filter analytically and showed that they vanish geometrically in a way that most recent values have more importance than older values. The weights are crucial because they determine how

[3] It is important to take into account that this parametric method requires sufficient additional structure to ensure identification. See Stock and Watson (2004).

the new information is incorporated into the estimation of the state variable and, therefore, into the forecasts. Following Koopman and Harvey, the expression for the predicted state at time $t+1$ given the information at time $t$ in our case is:

$$f_{t+1|t} = FK_t y_t + \sum_{j=1}^{t-1} FK_j y_j \prod_{i=j+1}^{t} F(I - K_i H) \tag{3.4}$$

where the $K_j$ is the Kalman gain at time $j$ and given by: $K_j = P_{j|j-1}H'(HP_{j|j-1}H' + R)^{-1}$. Notice that the Kalman gain depends inversely on the variance of the disturbances in the measurement equation $R$ and also on the predicted state variance $P_{j|j-1}$, which is influenced by the initial state variance $P_1$ and the variances of the disturbances in the transition equations $Q$.

The forecasts of the variables at time $t+1$ given the information at time $t$ are

$$y_{t+1|t} = Hf_{t+1|t} = HFK_t y_t + \sum_{j=1}^{t-1} HFK_j y_j \prod_{i=j+1}^{t} F(I - K_i H) \tag{3.5}$$

The first term of the summation is equivalent to the pooling term in Peña and Poncela (2004) and it is very important for the forecasts. The weight of this term depends on the signal-to-noise ratio ($Q/R$), the squared of the loading matrix ($H$) and whether the state variables are stationary or not ($F$).

The Kalman filter is designed in a way that it reaches the steady state in very few iterations and the Kalman gain will remain in its steady state value, which is constant. This creates problems in episodes of instability in variance. To understand why these models have problems to predict a recovery, suppose we are in the beginning of the recession and many variables drop dramatically. Assuming that the parameters are fixed or known, the forecast of the Kalman filter one period ahead will be driven mainly by the abnormal negative value of the last period, because past values and their weights are constant and neglectable. Over time, abnormal negative values will go to the second part of the sum, and even though their weights become smaller over time, they can still play an even more influential role than the most recent values. It is as if the model has lost memory and only the observations of the volatile period count. This is in agreement with Box-Steffensmeier and Lebo (2008) who demonstrated that methods like the Kalman filter or moving averages (i.e. rolling regressions or recursive regressions) in linear regression models cause problems with statistical inference and forecasting in presence of variance instability. They show that once the instability period is over, the model seems to lose its memory and does not use the past history any more. In other

words, the most volatile part dominates and the model has only short-run memory.

On the other hand, imagine that one variable or a set of variables with a small Kalman gain or loading coefficients are among the first to signal the recession. Given that their weights are going to be constant and small, these atypical values will hardly have an effect on the forecasts. If the weights of the Kalman filter were not constant but instead updated, they would react in some episodes in a way that the model would not lose memory. One possible solution is to make the loading matrix time-varying. Alternatively, we could modify the filter in order to allow the Kalman gain to change over time. This is essentially what we do in the presence of missing values where the gain is fixed to zero for those observations with the aim of giving them null weight.

In practice, what is going to happen in most cases is that the maximum likelihood estimation accommodates the parameters to reach maximum in-sample fit. This can cause problems in out-of-sample forecasting. Furthermore, there can arise convergence problems because the likelihood function is flatter or its shape is complicated, for instance because of multiple modes. Additionally, we could get negative or zero values for the estimated variance of idiosyncratic factors (Sentana, 2000). These so-called Heywood cases may be caused by including too many or too few common factors, $N$ and $T$ being too small to provide stable estimates, a misspecified model, etc. The incidence of Heywood cases increases with the variance of the idiosyncratic components and the maximum likelihood method is especially vulnerable.

### 3.2. Proposed solution

One possible way of dealing with instability of second moments is to introduce time-varying parameters in the loadings matrix. However, in a model of unobserved components it would create further identification problems and some additional assumptions would be required. Alternatively, we could consider switching regimes in the common factor for the first and/or second moments (e.g. Camacho et al. 2012), but it might be difficult to determine the number of regimes and the pattern of switching. Instead we propose a more parsimonious solution that involves modeling the conditional variance of the innovations of the common factor $Q_t$. Introducing heteroscedasticity in factor models may improve forecasting and statistical inference because it is going to affect the signal-to-noise ratio which is crucial for forecasting and to identify the components. We will assume that $Q_t$ follows either a GARCH or an Autoregressive Stochastic Volatility (ARSV) model. In this way the DFM implies the following conditional variance-covariance decomposition:

$$var_{t-1}(y_t) = HQ_tH' + R \tag{3.6}$$

but if the factors are covariance stationary (i.e. $E(Q_t) = Q$), we will have the same decomposition in unconditional terms:

$$E(\Sigma_t) = HE(Q_t)H' + R = HQH' + R \tag{3.7}$$

According to Sentana (1998) introducing heteroscedasticity in the common factors has interesting implications for the conditional correlation between two variables, which is given by:

$$\rho_{12t} = \frac{h_1 h_2 q_t}{\sqrt{(h_1^2 q_t + r_1)(h_2^2 q_t + r_2)}} \tag{3.8}$$

Introducing time-varying variance in the common factor $q_t$ captures the aforementioned stylized facts: the volatility clustering, the commonality in volatility clustering and the relationship between variance and correlation. This is, that periods when the variables are more correlated coincide with those when the variance of the variables increase simultaneously. Notice that the correlation is strongly related to the signal-to-noise ratios $q_t/r_1$ and $q_t/r_2$.

# 4. Monte Carlo experiment

## 4.1. Factor GARCH

The first experiment we have performed is to simulate 100 times a DFM for four series (sample size 200) with only one common factor. The common factor and the idiosyncratic components follow AR(2) processes, but our results are not sensitive to this choice.The innovations of the common factor are conditionally heteroscedastic. In particular, they follow a GARCH(1,1):[4]

$$Q_t = (1 - \alpha - \beta) + \alpha u_{t-1|t-1}^2 + \beta Q_{t-1} \tag{4.1}$$

We distinguish two cases: a) $\alpha = 0.3$,$\beta = 0.5$: persistent but smooth GARCH. b) $\alpha = 0.5$,$\beta = 0.3$: persistent but volatile conditional variance. With the simulated series

[4]Notice that the model assumes that the unconditional variance is 1. As we use variance targeting proposed by Engle and Mezrich (1996), the conditional variance is expressed in terms of the unconditional variance.

we simultaneously estimate the parameters[5] and the conditional variances by ML. This increases the efficiency of the estimation, especially in large samples, although it is computationally more difficult. Alternatively, we could use a two-step procedure, but Engle and Sheppard (2001) concluded that these estimators are not fully efficient as they use limited information. Notice that in the estimation step we take the number of factors and the lags of the idiosyncratic errors and factors as known, although this is in practice an additional source of error.

Our interest is to study the one-step-ahead out-of-sample forecasting properties of factor GARCH when we compare it with a standard homoscedastic DFM. The results of the simulations are collected in Table 1. What we theoretically expect is that an incorrectly specified model such as the homoscedastic one does a bad job. In contrasts, what we obtain according to the root mean squared error (RMSE) is that on average it performs comparable with the factor GARCH. Possibly the reason for the similar forecasting abilities is that GARCH introduces an estimation bias because the GARCH parameters are not very precisely estimated. As shown by Fiorentini and Sentana (2001), $\beta$ is much more imprecisely estimated than $\alpha$. Apart from this reason, Bos and Koopman (2004) and Harvey et al. (1992) demonstrated that in this situation the Kalman filter is no more optimal because the filter is evaluating a likelihood that is not linear any more as some of the parameters depend on squared observations. So, in reality it is evaluating a quasi-likelihood function. Furthermore, the estimation of GARCH models is very sensitive to the existence of non-normal residuals, outliers and structural breaks.

We also compare the factor GARCH with the DFM including stochastic volatility that will be studied in the next subsection. The model with stochastic volatility performs worse. Moreover, the test of forecast accuracy developed by Clark and West (2007) does not reject in any case the null hypothesis that the forecast accuracy of the heteroscedastic models is similar to the one in the homoscedastic case.

## 4.2. Factor stochastic volatility

The shocks governing the volatility may not necessarily be the innovations of the common factor. Therefore, a natural step is to extend the model to account endogenously for other types of shocks. This is the case of the ARSV model. Empirically it has been found that a simple ARSV fits the data equally well as more heavily parameterized GARCH models. Apart from that, ARSV is more flexible than GARCH models, even though it requires simulation methods to estimate the unobserved innovation of the variances.

[5] To guarantee positive variances we reparametrise the GARCH parameters in the following way: $\alpha = sin^2(\alpha^*)$ and $\beta = sin^2(\beta^*)(1-\alpha)$. The initial value for $Q_t$ is $Q_1 = E(Q_t)$.

The complexity of estimating these models introduces an additional estimation bias and the uncertainty in the estimation of the stochastic volatility must be taken into account inside the likelihood. This together with the fact that the model is no more linear and Gaussian gives as a result that there is no analytical expression for the likelihood, and thus, numerical methods are required to compute it. Regarding these methods, there are two approaches in the literature: the Monte Carlo Markov Chain (MCMC)[6] and the Sequential Monte Carlo (SMC)[7] simulation-based methods.

MCMC is an iterative algorithm and typically delivers smoothed estimates (i.e. smoothing algorithm as the estimations of the parameters and variables of interest are based on all the information available in the whole data set). In contrast, the SMC is a recursive algorithm which is more appropriate for real time or on-line analysis[8] (i.e. filtering algorithm). As noticed in Primiceri (2005), smoothed estimates are more efficient and suitable when our objective is to identify and estimate the evolution of unobservable states over time (e.g. identify structural shocks), whereas filtered estimates are better for forecasting. As the latter applies to our study, we consider simulation-based filtering SMC methods such as particle filters, which are also much easier to implement than MCMC methods.

The purpose of SMC methods is to sequentially update samples from posterior distributions via importance sampling and resampling techniques. According to Doucet, de Freitas and Gordon (2001) particle filters produce Monte Carlo approximations to posterior distributions by propagating simulated samples whose weights are updated against incoming observations and taking advantage of the state-space representations of dynamic models. Therefore, each particle is a sampled value of the state vectors and/or the parameters of interest. In the state-space framework, these filtering algorithms perform reasonably well at filtering states in non-linear and/or non-gaussian models. Actually Fernández-Villaverde and Rubio-Ramírez (2007) show how particle filtering is useful to estimate dynamic macroeconomic models, in particular, dynamic stochastic general equilibrium (DSGE) models because the economies can be non-linear and/or non-normal. Moreover particle filtering is a likelihood-based approach comparable to the Bayesian averaging, where the weights depend on the likelihood of each particle.

The most popular SMC filter is the Auxiliar Particle Filter (APF) of Pitt and Shephard (1999) that uses the optimal importance density to compute the importance weights in

[6] See Robert and Casella (2004) for an overview of MCMC.

[7] See Doucet, De Freitas and Gordon (2001) and Creal (2009) for a survey of SMC methods, and Fernández-Villaverde and Rubio-Ramírez (2004) for an illustrative example.

[8] Notice that the filtered estimates of the volatilities in DFM-SV estimated by SMC are directly comparable to the estimated volatilities in the DFM-GARCH because both rely on past information.

the sampling step. The optimality involves perfect adaptation of the algorithm. A similar but more efficient method is the Rao-Blackwellised Particle Filter (RBPF henceforth) (Chen and Liu, 2000; Andrieu and Docet, 2002). It is based on marginalization via Kalman Filter to reduce the Monte Carlo variation and improve numerical efficiency.

It is typically assumed that the parameters of the model (e.g. the loadings or the autoregressive parameters for the state factor in our DFMs) are known or given. In practice, however, these parameters are unknown and have to be estimated. This causes problems in the previous algorithms because they should sequentially update the parameters given the estimated filtered states, and viceversa, update the estimated filtered states given the estimated parameters. There are very few works that have dealt with this problem. One possible solution proposed by Aguilar and West (2000) is to estimate the parameters and the states with MCMC methods (e.g. Gibbs sampling) to fit the model to historical data, and then to carry out sequential particle filtering only on the states to forecast given the parameters previously estimated by MCMC. The problem is that MCMC usually is computationally very intensive. Therefore, running MCMC each time a forecast exercise is performed is not very practical, especially for real time analysis. What is necessary is a method capable of modifying the filtered and predicted values fast and efficiently as new information arrives. In this line of reasoning, a few works have adapted the sequential filtering algorithms to allow for sequential parameter learning. This learning process is based on one idea introduced by Gordon et al. (1993) in a different context and consists in adding small random perturbances to the parameter draws. This is a way of introducing an artificial evolution to the parameter as if they were time-varying even though they are constant over time. In reality, what is changing is the estimation of the parameters given the states. As the introduction of these shocks can lead to problems in the precision of the inferences, Liu and West (2001) considered a kernel smoothing of the parameters. They impose a Gaussian kernel with a shrinkage rule for the mean value (kernel location) to reduce over-dispersion, and a scale of the kernel that is a function of the smoothing parameter.

The aforementioned work of Liu and West (2001) proposed a general algorithm that incorporates this kernel parameter learning into the APF. However, in this paper we instead use the RBPF, also known as *mixture Kalman Filter*. It is preferable because it combines the standard and popular Kalman filter with Gaussian mixtures. Given the characteristics of our state-space model (i.e. conditionally linear and Gaussian), it is more efficient and flexible to use this extended version of the Kalman filter because it easily accommodates departures from normality and non-linearities. More details on this algorithm can be found in the Appendix.

Nevertheless, it should be noticed that there is evidence showing that the APF, and to a lesser extent RBPF, could degenerate for sequential parameter learning and result in inaccuracies, especially for the variance of the innovations of the volatility. This parameter is crucial, not only for particle filtering but also for MCMC. As Liu and West (2001) commented: "Sequential simulation-based filtering methods must always be combined with some form of periodic recalibration based on off-line analysis performed with much more computational time available than the filtering methods are designed to accommodate". They propose to monitor the learning process of the parameters and compare it with their values when the model is estimated by MCMC. But as in some cases the problem of parameter degeneracy is very serious, they also suggested to use the parameter values obtained by MCMC or Maximum Likelihood (also known as off-line methods) to avoid inaccuracies.

Additionally, when the number of parameters is very high, this problem could become so important that it leads to a *sample impoverishment* or *depletion*. The reason is that due to the high variance of the importance weights over time, very few particles are used in each iteration to approximate the posterior distribution. Because of that, it is necessary to monitor that there is no weight degeneracy or that the number of dead particles is not very high. In order to do that we compute several measures such as: i) the *survival rate*, $SR_t = (1 - N_t)/D$, where $N_t$ is the number of non-selected particles at time $t$ and $D$ the number of total particles; ii) the *effective sample size* (Liu, 1996), $ESSt = (1/\sum_{i=1}^{D} w_t^{(i)2})^{-1}$ where $w_t^{(i)}$ is the importance weight of particle $i$ at time $t$; and iii) the *Shannon entropy*, $SE_t = -\sum_{i=1}^{D} w_t^{(i)} ln(w_t^{(i)})$. We have to check for instance that the survival rate or the entropy do not decrease over time, and the effective sample size is not lower than 60-80% of the total particles generated. In practice, the performance of the parameter learning requires a bit of tuning of the width of the kernel and the variance of the artificial noise.

Now we repeat the same forecasting experiment performed in the previous subsection. First, we simulate 100 times a DFM for four series (sample size 200) and one common factor. But this time the innovations of the factor follow the SV model

$$ln(Q_t) = ln(Q_{t-1}) + w_t \tag{4.2}$$

Then, we estimate again 3 models: the homoscedastic DFM, the DFM with factor GARCH, and the DFM with factor SV[9]. The results are displayed in Table 2. In con-

[9]The resampling method used is the stratified sampling, and the discount factor in the parameter learning algorithm is 0.9. Different resampling methods and values of the discount factor led to very similar results.

trast to the GARCH case, this time the SV clearly outperforms the GARCH and the homoscedastic models. In fact, the test of forecast accuracy does not support the hypothesis that the homoscedastic model performs similar at forecasting. Nevertheless, although it is not shown, notice that in the SV model the parameters are estimated with a small bias due to the kernel smoothing method.

## 5. Empirical application: Stock and Watson's (1991) DFM

Stock and Watson (SW, 1991) proposed a simple DFM to estimate a coincident indicator of the economic activity for the United States from four monthly series. These series are carefully monitored by the NBER in order to date a chronology of the peaks and troughs, or beginning of recessions and expansions. This model became very popular to construct business cycle indicators and has subsequently been extended by Kim and Nelson (1999) and Mariano and Murasawa (2003). However, Stock and Watson (2008) showed in their work about forecasting with DFM in presence of instabilities that the estimates of the common factor are quite stable and, in contrast, the estimates of the loadings or regression coefficients are quite unstable. Therefore, they concluded that the best strategy to produce accurate forecasts in this framework is to use estimates of the common factors using the full sample, and estimates of the loadings using only a subsample or time-varying estimations. This is related to the idea that the best predictors are not always the same indicators or, in other words, the time-varying relationship among variables. Later on many subsequent papers considered a large number of indicators because large cross-sections provide insurance against structural instabilities.

Using the same specification as SW(1991) and our proposed extensions, we perform next a forecasting exercise out-of-sample and in pseudo real time. We predict one-period ahead from 1994.01 until 2011.04, re-estimating the model each period but using the last vintage of data available (May 2011). The results are collected in Table 3.

We see that introducing stochastic volatility in the common factor reduces the one-step-ahead mean squared error (MSE) and the mean absolute error (MAE) when we consider the whole period, 1994.01-2011.04, and especially for the great recession, 2007.12-2009.06. In contrast, introducing GARCH does not lead to significant forecasting improvements.

We also compute forecast accuracy tests to compare all the specifications with the homoscedastic DFM. There are two well-known one-sided tests in the literature: Diebold and Mariano (1995) and Clark and West (2007). The former is very popular because it is very robust to non-quadratic loss functions and when the forecast errors are non-gaussian or have non-zero mean or under serial and contemporaneous correlation. However, it is

more suitable to test forecast accuracy in non-nested models and it does not take into account the noise introduced by the estimation of parameters. In order to do that the second test introduces a correction term. Apart from that, Clark and West's test is designed to forecast evaluation in nested models, as in our case. It compares a small model with a larger one which encompasses it. The results of Clark and West's test are collected on Table 4. Notice that to compute it we assume a quadratic loss function and asymptotic normality of the computed statistic, and we use Barlett's window to compute the long-run variance together with the optimal lag truncation parameter suggested by Newey and West (1994).

Considering both the MSE and the statistical test, we can conclude that in general the models with heteroscedasticity in the common factor improve the forecast accuracy for most of the series and also during the great recession. However, as stressed by Diebold and Mariano, it is important to consider that the superiority of a particular model in terms of forecasts accuracy does not necessarily imply that forecasts from other models contain no additional information. Therefore, our conclusion does not mean that the forecasts of the homoscedastic DFM are wrong or not informative at all. But in specific situations such as in a very serious crisis like the great recession, or when the level of uncertainty is very high, it could be worth to consider heteroscedastic models.

As we are going to see next, there are some additional advantages of our approach. First, it is straightforward to obtain the implicit correlations. For simplicity we have computed once more our measure of comovement with the estimated correlations that is directly comparable with the recursive measure computed before.

In Figure 3 we see that in all cases we observe an inverse S shape. This means that the correlations would have been higher in the beginning and would have gone down with the great moderation. Nevertheless, there are some symptoms of increasing correlation in the last part of the sample. This is in stark contrast with the constant correlation assumption in the standard DFMs.

As a useful by-product of our approach we get an estimate of the volatility of the common factor, which in this business cycle model can be interpreted as a measure of the broad macroeconomic risk. The GARCH and SV (filtered) estimates of the common factor's volatility are displayed in Figure 4. It picks up the two facts mentioned before: the great moderation and the higher volatility during the recessions. Notice that both measures are realized volatilities as they are filtered estimations.

We have performed the exercise of estimating the smoothed volatility in the DFM with stochastic volatility in the common factor by MCMC methods which is displayed in Figure 5. We observe the same pattern as before in the filtered volatility but in an even

clearer way.

One advantage more of our SMC method is that as it is designed for non-linear or non-gaussian models, it is straightforward to introduce additional non-linearities in our DFM-SV. In this way, we have estimated the model considering Markov Switching in the mean of the factor, with two states corresponding to expansion and recession. It is also straightforward to obtain the probability of being in recession as it is shown in Figure 6. The periods when the probabilities of recession are near one coincide very closely with the official recessions dated by the NBER (shaded areas). We also observe that the model has more difficulties to signal the exit of the recessive phase in the last 3 recessions, characterized by jobless recoveries.

## 6. Conclusions

Episodes such as the recent Great Recession and moments of especially high uncertainty have underlined the necessity of considering the changing level of uncertainty in models for forecasting in macroeconomics such as the DFM. In this work we have extended the standard DFM in a parsimonious way to take into account time-varying correlations and variances. We do this by introducing heteroscedasticity (i.e. GARCH or ARSV) in the common factor. We have also proposed a sequential Monte Carlo method to estimate stochastic volatility in DFM, easier to implement than MCMC and, therefore, more appropriate for performing forecasting exercises. Additionally, this method is fast, efficient and robust to non-linearities and departures from gaussianity. Furthermore, we have shown that the heteroscedastic models have better forecasting properties at short-run (one period ahead) than homoscedastic models, especially in these specific episodes. Apart from that, it takes the time-varying correlations between variables into account and delivers the volatility of the common factor which we can interpret as an indicator of global macroeconomic risk. Finally, it is possible to extend these models for additional non-linearities such as markov switching in the mean of the factor to compute probabilities of being in recessions or expansions.

Regarding our proposed method to introduce SV in DFM, it has however some shortcomings that must be considered. There could arise problems with the parameter learning, especially when the number of parameters is very high, the learning process is too slow or the priors are not very realistic. The problem is that how to choose effective particles still lacks of rigorous justification. And in some cases it could happen that the Monte Carlo error grows exponentially. Although there are some recent proposals in the literature to improve this methodology, there is no agreement on which is the best one.

Most of them try to improve SMC by introducing MCMC steps. This requires further research.

One step ahead in our research agenda is to extend these models to real-time data. These datasets have interesting characteristics: the data have mixed frequencies, missing observations, and ragged-ends, because the indicators are not released at the same time. And therefore, the level of uncertainty is higher and the variance problems are even more relevant. Nevertheless, to adapt the SMC methods is complicated because the number of missing data is crucial. When the rate of missing data increases, it is harder to achieve a certain tolerance and a large number of particles is needed.

## References


[1] O. Aguilar and M. West. Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics*, 18:338–357, 2000.

[2] L. Alessi, M. Barigozzi, and M. Capasso. Estimation and forecasting in large datasets with conditionally heteroskedastic dynamic common factors. *ECB working paper*, 1115, 2009.

[3] C. Andrieu and A. Doucet. Particle filtering for partially observed gaussian state space models. *Journal of Royal Statistical Society*, 64, 4:827–836, 2002.

[4] E. Angelini, G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler. Short-term forecasts of euro area gdp growth. *Econometrics Journal*, 14:c25–c44, 2011.

[5] L. Bauwens, S. Laurent, and J.V. Rombouts. Multivariate garch models: a survey. *Journal of Applied Econometrics*, 21, 1:79–109, 2006.

[6] T. Bollerslev. Modelling the coherence in short-run nominal exchange rates: A multivariate generalised arch model. *Review of Economics and Statistics*, 72:498–505, 1990.

[7] C.S. Bos and S.J. Koopman. State space models with a common stochastic variance. *Journal of Business and Economic Statistics*, 22:346–357, 2004.

[8] C.S. Bos and S.J. Koopman. Models with time-varying mean and variance: a robust analysis of us industrial production. *Tinbergen Institute Discussion Paper TI*, 2010-017/4, 2010.

[9] J. M. Box-Steffensmeier and M. Lebo. Dynamic conditional correlations in political science. *Journal of Political Science*, 52, 3:688–704, 2008.

[10] A.F. Burns and W.C. Mitchell. Measuring business cycles. *NBER Book Series Studies in Business Cycles*, 1947.

[11] M. Camacho and G. Pérez-Quirós. Introducing the euro-sting: Short-term indicator of euro area growth. *Journal of Applied Econometrics*, 25, 4:663–694, 2010.

[12] M. Camacho, G. Pérez-Quirós, and P. Poncela. Markov-switching dynamic factor models in real time. *Banco de España Working paper*, 1205, 2012.

[13] L. Cappielo, R. F. Engle, and K. Sheppard. Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics*, 4, 4:537–572, 2006.

[14] M. Chauvet and S. Potter. Business cycle monitoring with structural changes. *International Journal of Forecasting*, 26, 4:777–793, 2010.

[15] R. Chen and J.S. Liu. Mixture kalman filters. *Journal of Royal Statistical Society*, 62, 3:493–508, 2000.

[16] T.E. Clark and K.D. West. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138:291–331, 2007.

[17] D. Creal, S.J. Koopman, and E. Zivot. Extracting a robust u.s. business cycle using a time-varying multivariate model-based bandpass filter. *Journal of Applied Econometrics*, 25:695–719, 2010.

[18] D. D. Creal. A survey of sequential monte carlo methods for economics and finance. *Working paper, Department of Econometrics, Vrije Universiteit Amsterdam*, 2009.

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

[20] F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–263, 1995.

[21] R. Douc, O. Cappé, and E. Moulines. Comparison of resampling schemes for particle filtering. . *In 4th International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2005.

[22] A. Doucet, N. de Freitas, and N. Gordon. Sequential monte carlo methods in practice. 2001.

[23] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50, 4:987–1007, 1982.

[24] R. F. Engle. Anticipating correlations: A new paradigm for risk management. 2009.

[25] R. F. Engle and J. Mezrich. Garch for groups. *Econometrica*, 9, 8:36–40, 1996.

[26] R. F. Engle and K. Sheppard. Theoretical and empirical properties of dynamic conditional correlation multivariate garch. *NBER Working Papers*, 8554, 2001.

[27] A. Espasa, S. Pellegrini, and E. Ruiz. Prediction intervals in conditionally heteroscedastic time series with stochastic components. *International Journal of Forecasting*, 27, 2:308–319, 2010.

[28] B. T. Ewing and M.A. Thompson. Industrial production, volatility, and the supply chain. *International Journal of Production Economics*, 115:553–558, 2008.

[29] W. Fang, S. M. Miller, and C. Lee. Cross-country evidence on output-growth volatility nonstationary variante and garch models. *Scottish Journal of Political Economy*, 55, 4, 2008.

[30] J. Fernández-Villaverde and J. F. Rubio-Ramírez. Sequential monte carlo filtering: an example. *Mimeo*, 2004.

[31] J. Fernández-Villaverde and J. F. Rubio-Ramírez. Estimating macroeconomic models: A likelihood approach. *Review of Economic Studies*, 74:1059–1087, 2007.

[32] G. Fiorentini and E. Sentana. Identification, estimation and testing conditionally heteroskedastic factor models. *Journal of Econometrics*, 102:143–164, 2001.

[33] R. Gerlach, C. Carter, and R. Kohn. Efficient bayesian inference for dynamic mixture models. *Journal of American Statistical Association*, 95, 451:819–828, 2000.

[34] J. Geweke. The dynamic factor analysis of economic time series. *Latent Variables in Socio-Economic Models*, 1977.

[35] P. Giordani and R. Kohn. Efficient bayesian inference for multiple change-point and mixture innovation models. *Journal of Business and Economic Statistics*, 26, 1:66–77, 2008.

[36] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F on Radar and Signal Processing*, 140, 2:107–113, 1993.

[37] J. Hamilton. Macroeconomics and arch. *NBER Working Paper series*, 14151, 2008.

[38] A. Harvey, E. Ruiz, and E. Sentana. Unobserved component time series models with arch disturbances. *Journal of Econometrics*, 52, 1-2:129–157, 1992.

[39] K. Y. Ho, A. K. Tsui, and Z. Zhang. Modelling volatility asymmetry of business cycles in the u.s. *Mathematics and Computers in Simulation*, 79:2856–2868, 2009.

[40] K. Y. Ho and A. K. C. Tsui. Asymmetric volatility of real gdp: some evidence from canada, japan, the united kingdom, and the united states. *Japan and the World Economy*, 15, 4:437–445, 2003.

[41] C.J. Kim and C.R. Nelson. State-space models with regime switching. 1999.

[42] G. Koop, R. Leon-Gonzalez, and R.W. Strachan. On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, 33, 4:997–1017, 2009.

[43] S.J. Koopman and A. Harvey. Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control*, 27:1317–1333, 2003.

[44] J. Liu and M. West. Combined parameter and state estimation in simulation based filtering. *Sequential Monte Carlo Methods in Practice*, 2001.

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

[46] D. B. Nelson and D. P. Foster. Asymptotic filtering theory for univariate arch models. *Econometrica*, 62:1–41, 1994.

[47] W. K. Newey and K. D. West. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61, 4:631–654, 1994.

[48] D. Peña and P. Poncela. Forecasting with nonstationary dynamic factor models. *Journal of Econometrics*, 119:291–321, 2004.

[49] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94, 446:590–599, 1999.

[50] G. Primiceri. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72, 3, 2005.

[51] C.P. Robert and G. Casella. Monte carlo statistical methods. 2004.

[52] T. J. Sargent and C. Sims. Business cycle modelling without pretending to have too much a priori economic theory. *New Methods in Business Cycle Research*, 1977.

[53] E. Sentana. The relation between conditionally heteroskedastic factor models and factor garch models. *The econometric Journal*, 1:1–19, 1998.

[54] E. Sentana. The likelihood function fo conditionally heteroskedastic factor models. *Annales d'Économie et de Statistique*, 58:1–19, 2000.

[55] J. H. Stock and M. W. Watson. Median unbiased estimation of coefficient variance in a time-varying parameter model. *Journal of the American Statistical Association*, 93, 441:349–358, 1998.

[56] J. H. Stock and M.W. Watson. Forecasting with many predictors. *Handbook of Economic Forecasting*, 2004.

[57] J. H. Stock and M.W. Watson. Why has u.s. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39, 1:13–33, 2007.

[58] J. H. Stock and M.W. Watson. Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*, 2008.

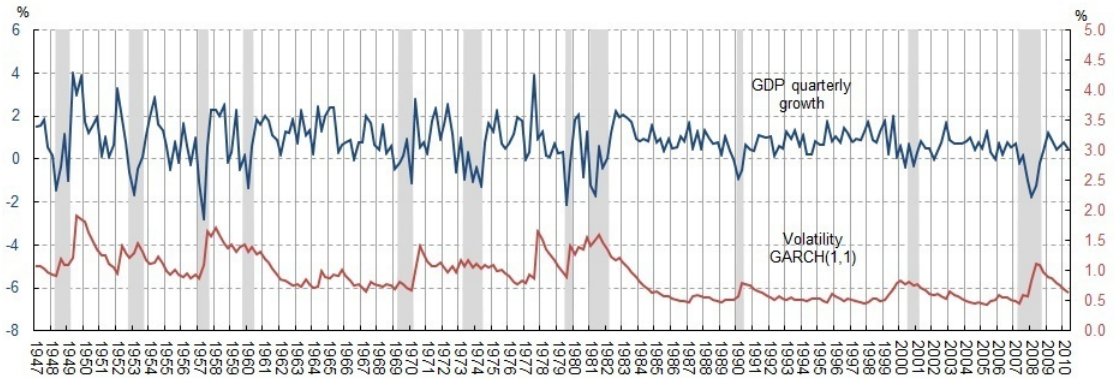
[59] J. H. Stock and M.W. Watson. The evolution of national and regional factors in us housing construction. *Volatility and time series econometrics: essays in honor of Robert F. Engle*, 2010.

[60] A.A. Weiss. Arma models with arch errors. *Journal of Time Series Analysis*, 5:129–143, 1984.

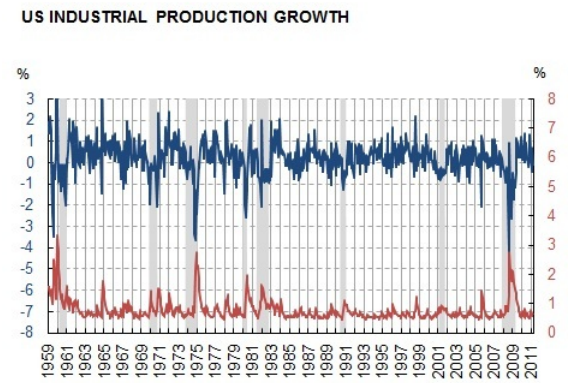
[61] M. West. Approximating posterior distributions by mixtures. *Journal of the Royal Statistical Society (Ser. B)*, 55, 2:409–422, 1993.

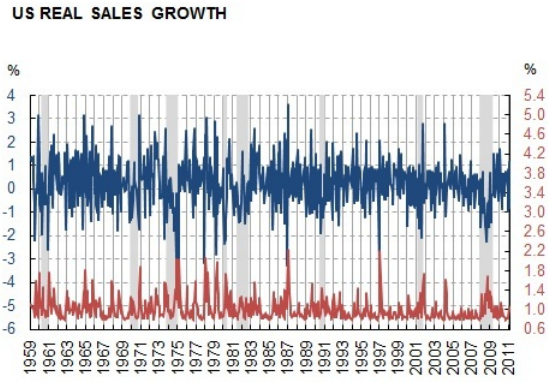
# Figures and Tables

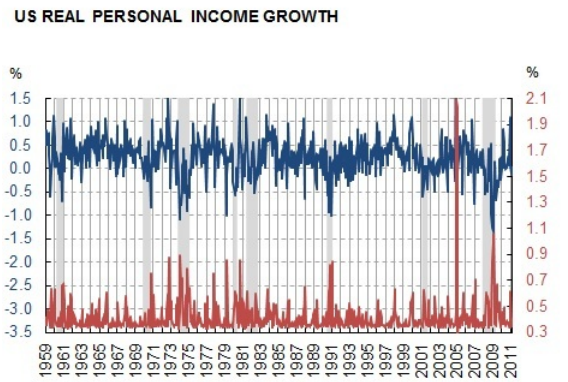
Figure 6.1: Volatility of some macroeconomic variables in the US

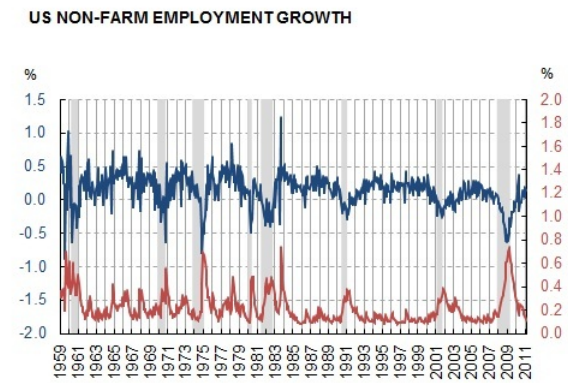



(a) Real GDP



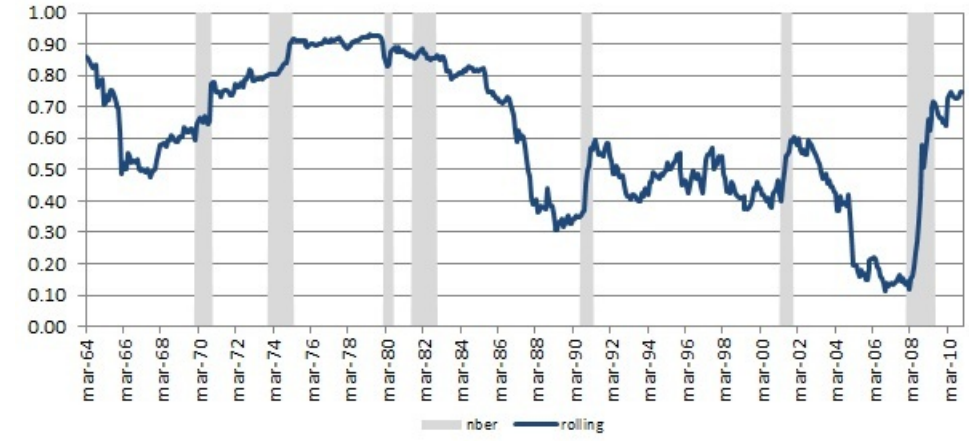



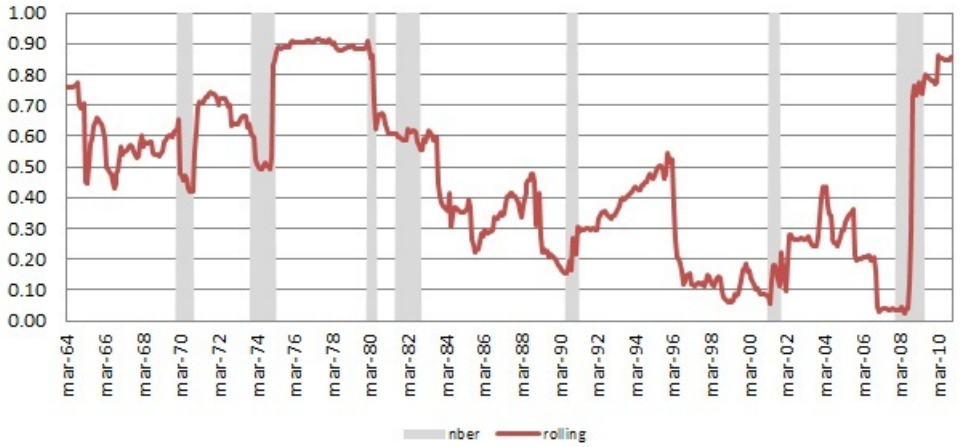




(b) Several monthly indicators

Figure 6.2: Measures of comovement and volatility comovement




(a) Comovement




(b) Volatility comovement

Figure 6.3: Dynamic correlations

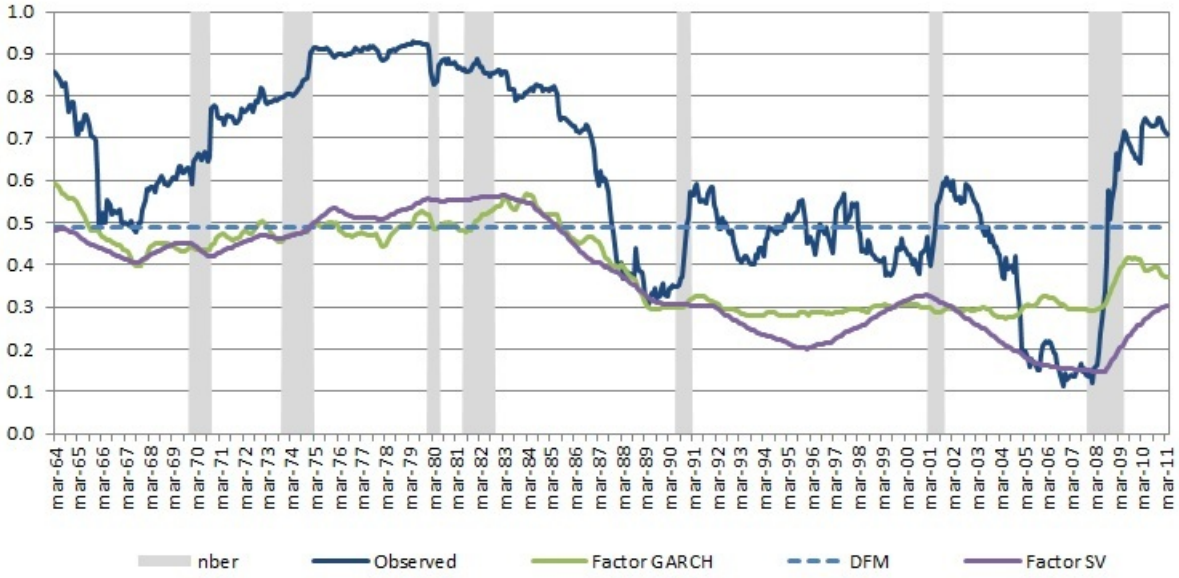


Figure 6.4: Common factor conditional variance. Filtered values

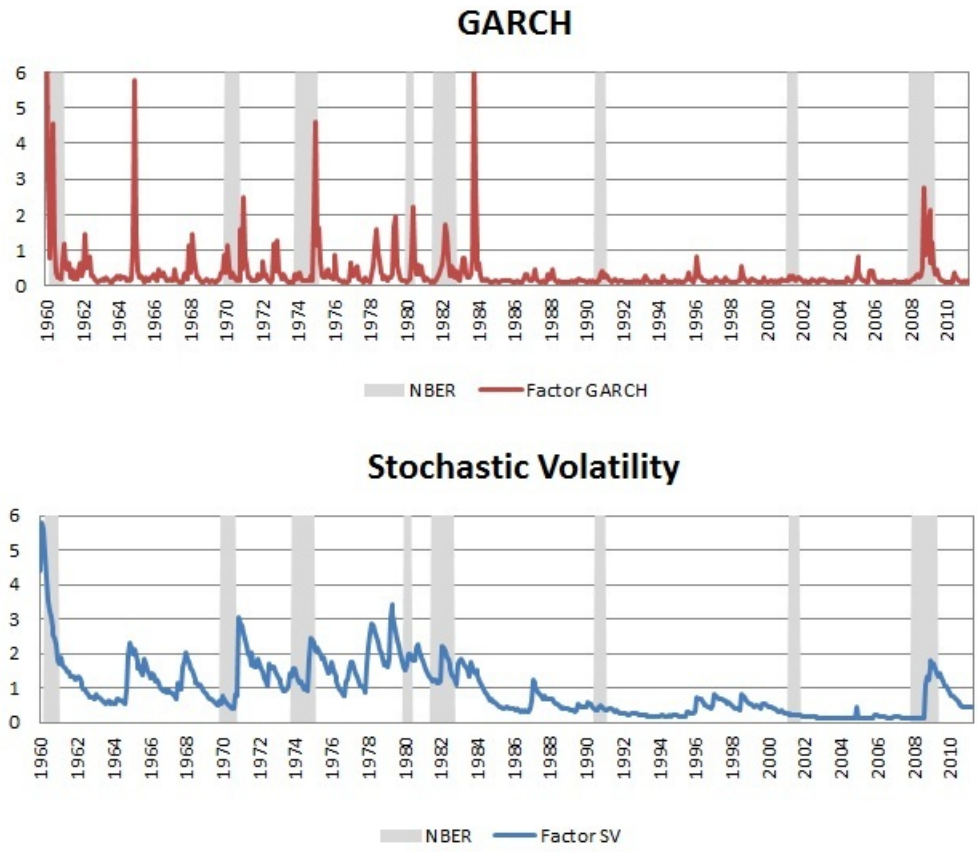


Figure 6.5: Common factor conditional variance. Smoothed values

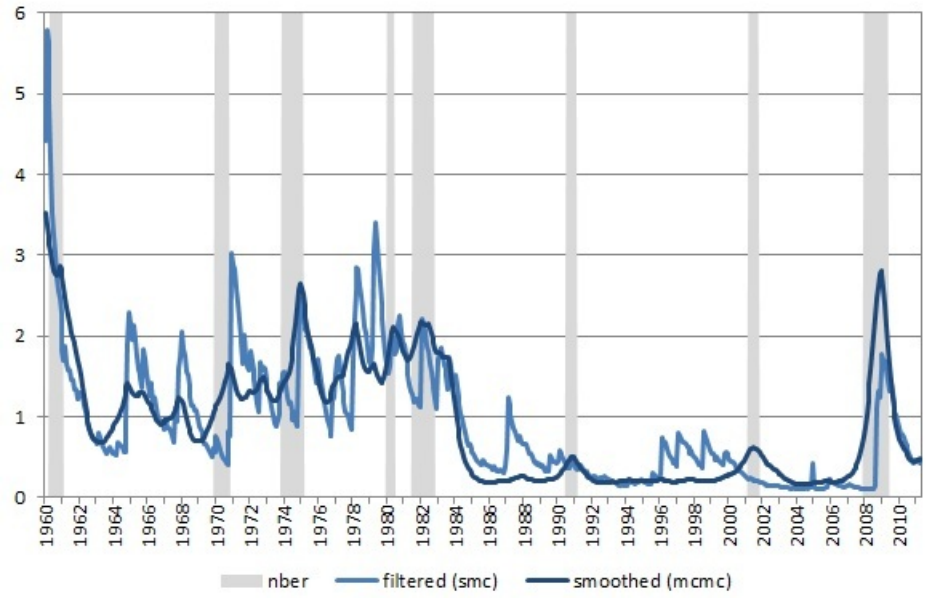



Figure 6.6: Filtered probabilities of recession

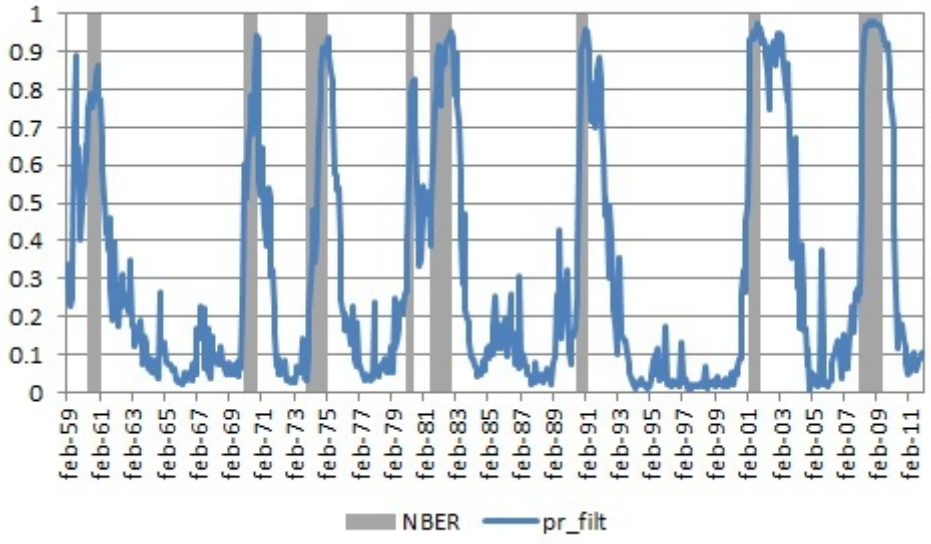


Table 1: Simulation exercise: Factor GARCH

RMSE one-period ahead

| Estimated models | variable 1 | variable 2 | variable 3 | variable 4 | average |
|---|---|---|---|---|---|
| Homosc. DFM | 0.65 | 0.61 | 1.14 | 1.10 | 0.88 |
| Factor GARCH (real) | 0.63 | 0.61 | 1.11 | 1.08 | 0.86 |
| Factor SV | 1.08 | 1.00 | 1.28 | 1.33 | 1.17 |

Note: ** corresponds to the significance level of 5% of the Clark and West (2005) forecast accuracy test.

(a) Persistent GARCH

RMSE one-period ahead

| Estimated models | variable 1 | variable 2 | variable 3 | variable 4 | average |
|---|---|---|---|---|---|
| Homosc. DFM | 0.71 | 0.66 | 1.15 | 1.07 | 0.90 |
| Factor GARCH (real) | 0.70 | 0.66 | 1.14 | 1.05 | 0.89 |
| Factor SV | 1.13 | 1.03 | 1.36 | 1.34 | 1.21 |

Note: ** corresponds to the significance level of 5% of the Clark and West (2005) forecast accuracy test.

(b) Volatile GARCH

Table 2: Simulation exercise: Factor SV

RMSE one-period ahead

| Estimated models | variable 1 | variable 2 | variable 3 | variable 4 | average |
|---|---|---|---|---|---|
| Homosc. DFM | 2.25 | 2.37 | 2.59 | 2.72 | 2.48 |
| Factor GARCH | 2.13 | 2.20 | 2.46 | 2.58 | 2.34 |
| Factor SV (real) | 1.92** | 1.98** | 2.39** | 2.34** | 2.16** |

Note: ** corresponds to the significance level of 5% of the Clark and West (2005) forecast accuracy test.

Table 3: Forecasting exercise with Stock and Watson's (1991) DFM

| | IPI | | INC | | SALES | | EMP | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| Homosc. DFM | 0.93 | 0.67 | 0.46 | 0.43 | 1.54 | 0.95 | 0.02 | 0.12 | 0.74 | 0.54 |
| Factor GARCH | 0.90 | 0.65 | 0.45 | 0.43 | 1.51 | 0.93 | 0.02 | 0.12 | 0.72 | 0.53 |
| Factor SV | 0.49 | 0.50 | 0.28 | 0.33 | 0.79 | 0.67 | 0.03 | 0.15 | 0.40 | 0.41 |

(a) Period 1994.01-2011.05

| | IPI | | INC | | SALES | | EMP | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| Homosc. DFM | 3.94 | 1.51 | 0.59 | 0.61 | 1.71 | 1.03 | 0.07 | 0.20 | 1.58 | 0.84 |
| Factor GARCH | 3.93 | 1.51 | 0.57 | 0.59 | 1.75 | 1.06 | 0.07 | 0.20 | 1.58 | 0.84 |
| Factor SV | 1.96 | 1.00 | 0.37 | 0.49 | 1.02 | 0.82 | 0.10 | 0.30 | 0.86 | 0.65 |

(b) Great recession

Table 4: P-values from the Clark and West (2007) forecast accuracy test

| | IPI | INC | SALES | EMP |
|---|---|---|---|---|
| Factor GARCH | 0.25 | 0.15 | 0.00 | 0.01 |
| Factor SV | 0.00 | 0.01 | 0.00 | 0.00 |

(a) Period: 1994-2011

| | IPI | INC | SALES | EMP |
|---|---|---|---|---|
| Factor GARCH | 0.38 | 0.18 | 0.29 | 0.33 |
| Factor SV | 0.00 | 0.00 | 0.00 | 0.00 |

(b) Great recession

# Appendix

## A. Modified Kalman filter for DFM with GARCH

Here we detail the steps of the extended Kalman filter to account for time-varying variances in the factor innovations following a GARCH model. First, the state-space representation of this model is:

$$y_t = \pi + Hf_t + \xi_t \tag{A.1}$$

$$f_t = Ff_t + u_t \tag{A.2}$$

$$Q_t = (1 - \alpha - \beta) + \alpha(u^2_{t-1|t-1} + P_{t-1|t-1}) + \beta Q_{t-1} \tag{A.3}$$

And the steps are:

1. Initialise $f_{0|0}$, $P_{0|0}$ and $Q_{0|0}$.
2. For $t = 1, 2, .., T$

   a) In the forecasting step the states and their variances together with the volatilities of the factor innovations (and idiosyncratic innovations, if we also consider this case) at time $t$ are estimated using the information available until $t-1$ with the next equations:

   $$f_{t|t-1} = Ff_{t-1|t-1} \tag{A.4}$$

   $$P_{t|t-1} = FP_{t-1|t-1}F' + Q_{t|t-1} \tag{A.5}$$

   $$Q_{t|t-1} = (1 - \alpha - \beta) + \alpha(u^2_{t-1|t-1} + P_{t-1|t-1}) + \beta Q_{t-1|t-1} \tag{A.6}$$

   Notice that the variance matrix of the states $P_{t-1|t-1}$is included in the second term in the right hand side of A.6. Harvey et al. (1992) introduce this correction term, given that the factor is an unobserved component and must be estimated, to take the uncertainty in the factor estimates into account.

   b) As soon as new data are available at time $t$, we compute the forecast errors and their corresponding variances,

   $$v_t = y_t - \pi - Hf_{t|t-1} \tag{A.7}$$

   $$\Sigma_t = HP_{t|t-1}H' + R \tag{A.8}$$

and update the estimates of the states together with their variances, and also the volatilities.

$$f_{t|t} = f_{t|t-1} + P_{t|t-1}H'\Sigma_t^{-1}v_t \tag{A.9}$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}H'\Sigma_t^{-1}HP_{t|t-1} \tag{A.10}$$

$$Q_{t|t} = Q_{t|t-1} \tag{A.11}$$

The log-likelihood function is calculated as function of the forecast errors and their variances.

$$loglik_t = -0.5(\ln(2\pi) + \ln(|\Sigma_t|) + v_t'\Sigma_t^{-1}v_t) \tag{A.12}$$

# B. Rao-Blackwellized Particle Filter (RBPF)

Rao-Blackwellized Particle Filter (RBPF) is an efficient sequential Monte Carlo (SMC) method because it recycles the simulated random variables by means of the popular Kalman filter algorithm. It is based on marginalisation and gaussian mixtures. In each iteration candidate values for the unknown variables (also called particles, because this methodology was developed in Physics) are generated. The name of Rao-Blackwellised is due to the Rao-Blackwellised theorem, which says that the expected value of any estimator conditioned on the information of a sufficient statistic is always better in terms of mean squared error than the estimator itself. In this case the proposed estimators given the information of the Kalman filter are better than the estimators themselves. Assuming that the values of all the parameters $\theta$(i.e. loadings, autoregressive coefficients, variances) are known, the purpose of the filter is to generate particles from a posterior density (or filtering density) $P(f_t|I_t, \theta)$ using the likelihood $P(y_t|I_t, \theta)$ and the prior density (or forecasting density) $P(f_t|I_{t-1}, \theta)$ in the nex way:

$$P(f_t|I_t, \theta) \propto P(y_t|f_t, \theta)P(f_t|I_{t-1}, \theta) \tag{B.1}$$

The observation equation in the DFM provides information about the likelihood. And given that the DFM is conditionally linear and gaussian, the likelihood is also normal and can be written as a function of the forecast errors $v_t$ and their variances $\Sigma_t$:

$$P(y_t|f_t, \theta) \sim N(v_t, \Sigma_t) \tag{B.2}$$

The state equation is useful to infer the prior distribution. Again, given that the model is conditionally linear and gaussian, this density will be normal with mean $Ff_{t-1|t-1}$ and

variance $FP_{t-1|t-1}F' + Q_{t-1}$, this is, A.4 and A.5. So, the prior density is

$$P(f_t|I_{t-1}, \theta) = P(f_t|f_{t-1}, \theta) \sim N(Ff_{t-1|t-1}, FP_{t-1|t-1}F' + Q_{t-1}) \tag{B.3}$$

The posterior density is therefore approximated in this discrete way:

$$P(f_t|I_t, \theta) \simeq \sum_{j=1}^{D} w_t^{(j)} P(f_t|f_{t-1}^{(j)}, \theta) \tag{B.4}$$

where $w_t^{(j)}$are the importance weights which are a function of the likelihood function. Thus, the posterior density is approximated by a mixture of normals $P(f_1|f_{t-1}^{(j)}, \theta)$ with weights $w_t^{(j)}$. Due to that and the use of Kalman filter, this method is also known as *Mixture Kalman filter*.

Once the philosophy is clear, next we explain the steps of the algorithm.

1. INITIALISATION: At time 0 we generate $D$ initial random particles for the states $\left\{f_{0|0}^{(j)}\right\}_{j=1}^{D}$. Their variances $\left\{P_{0|0}^{(j)}\right\}_{j=1}^{D}$, the volatilities of the factor innovations $\left\{Q_0^{(j)}\right\}_{j=1}^{D}$, and the importance weights $\left\{w_0^{(j)}\right\}_{j=1}^{D}$ are also initialised. Typically $f_{0|0}^{(j)}$ follows a standard normal, $P_{0|0}^{(j)}$ is the identity matrix, $Q_0^{(j)}$ is a random draw of a log-normal distribution and $w_0^{(j)} = 1/D$.

2. For $t = 1, 2, ..., T$

   a) PREDICTION step of the Kalman filter to get the state variables $\left\{f_{t|t-1}^{(j)}\right\}_{j=1}^{D}$ and their variances $\left\{P_{t|t-1}^{(j)}\right\}_{j=1}^{D}$. We also compute the forecast errors $\left\{v_t^{(j)}\right\}_{j=1}^{D}$ and their variances $\left\{\Sigma_t^{(j)}\right\}_{j=1}^{D}$, and the likelihood $\left\{lik_t^{(j)}\right\}_{j=1}^{D}$ given the observed $y_t$.

   b) Compute the IMPORTANCE WEIGHTS for each particle from the likelihood function $\left\{\tilde{w}_t^{(j)}\right\}_{j=1}^{D}$ and normalize them $w_t^{(j)} = \tilde{w}_t^{(j)} / \sum_{j=1}^{D} \tilde{w}_t^{(j)}$. This normalization is important for the next step.

   c) RESAMPLING of the particles: This step is necessary to reduce the sampling variability of the generated particles and to stabilise the algorithm. It consists in generating $D$ values of a multinomial $k$ which takes values $1, 2, ..., D$ with probabilities $w_t^{(j)}$, and select those particles for the states $\left\{f_{t-1|t-1}^{(k_j)}\right\}_{j=1}^{D}$and their variances $\left\{P_{t-1|t-1}^{(k_j)}\right\}_{j=1}^{D}$, and the volatilities of the

factor innovations $\left\{Q_{t-1}^{(k_j)}\right\}_{j=1}^{D}$. In the literature some small modifications have been proposed in order to reduce the resampling variance or Monte Carlo variation such as the stratified resampling (for more details see Douc, Cappé and Moulines (2005)).

d) UPDATING step of the Kalman filter to get the filtered values of the state variables $\left\{f_{t|t}^{(j)}\right\}_{j=1}^{D}$ and their variances $\left\{P_{t|t}^{(j)}\right\}_{j=1}^{D}$. We also obtain the updated volatilities of the factor innovations $\left\{Q_t^{(j)}\right\}_{j=1}^{D}$ given that $\ln Q_t^{(j)} \sim N(\ln Q_{t-1}^{(j)}, \sigma_W^2)$ and the weights $w_t^{(j)} = 1/D$.

## C. Parameter Kernel Smoothing (PKS)

Usually not only the state vector but the parameters are a priori unknown. This means that a new unkonwn term $\theta$ is included in the posterior density, in a way that it becomes a joint density of the state vector $f_t$ and the parameters $\theta$given the information until time $t$, $P(f_t, \theta|I_t)$. Applying Bayes' theorem,

$$P(f_t, \theta|I_t) \propto P(y_t|f_t, \theta)P(f_t, \theta|I_{t-1}) \propto P(y_t|f_t, \theta)P(f_t|f_{t-1}, \theta)P(\theta|I_{t-1}) \tag{C.1}$$

the joint posterior density $P(f_t, \theta|I_t)$ is proportional to the likelihood $P(y_t|f_t, \theta)$, the conditional or forecasting density of the state variable given the parameters $P(f_t|f_{t-1}, \theta)$, and the density of the parameters given the information until $t-1$, $P(\theta|I_{t-1})$. Under the assumption of known parameters, the latter density is degenerate and we can skip that last term and in the joint distribution in C.1. But more realistically if the parameters are unknown, the density of the parameters $P(\theta|I_{t-1})$ must be approximated to obtain draws from it. As already explained in section 4.2, one way of solving this issue is to treat the parameters as time varying, even though they are fixed, by adding small random disturbances to the parameters. Thus, the state vector is augmented with $\theta_t$. But it is important to clarify that $\theta_t$ means that our estimation about the values of the parameters changes with the information available. However, the parameters are actually fixed. Typically the following parameter learning evolution is imposed over the $D$ draws

$$\theta_t^{(j)} = \theta_{t-1}^{(j)} + \zeta_t^{(j)} \tag{C.2}$$

with $\zeta_t^{(j)} \sim N(0, W_t^{(j)})$ for $j = 1, 2, .., D$ and $t = 1, 2, ..., T$.

This artificial evolution could lead to very diffuse values for the draws of the parameters, and hence, cause problems of precision or loss of information. For this reason, West (1993) proposed to smooth these draws using kernel smoothing methods. In this way, we get draws of $\theta$ at time $t$ given the information until $t-1$ by means of the next discrete Monte Carlo approximation as a weighted mixture of normals

$$P(\theta_t|I_{t-1}) \approx \sum_{j=1}^{D} w_{t-1}^{(j)} N(\theta_t|m_{t-1}^{(j)}, h^2 V_{t-1}) \tag{C.3}$$

where $N(\bullet|m_{t-1}, h^2V_{t-1})$ is a multivariate normal density (i.e. Gaussian kernel) with mean $m_{t-1}$ (i.e. kernel location) and variance $h^2V_{t-1}$, and $w_{t-1}^{(j)}$ are the importance weights. Notice that $h$ is a smoothing parameter, strictly positive, and $V_t = \sum_{j=1}^{D}(\theta_{t-1}^{(j)} - \bar{\theta}_{t-1})^2/D$ is the Monte Carlo posterior variance and represents the kernel rotation and scaling. Furthermore, the next shrinkage rule for the mean $m_{t-1}^{(j)}$ is going to push draws of $\theta_t^{(j)}$towards the Monte Carlo finite mean $\bar{\theta}_{t-1} = \sum_{j=1}^{D} \theta_{t-1}^{(j)}/D$ and avoid over-dispersion

$$m_{t-1}^{(i)} = a\theta_{t-1}^{(j)} + (1-a)\bar{\theta}_{t-1} \tag{C.4}$$

with the number $a$ specified as function of a discount factor $\delta \in (0, 1]$, this is, $a = (3\delta - 1)/2\delta$. In practice $a$ typically takes values between 0.95 and 0.99. The smoothing parameter $h$ depends on $a$, usually specified in this way $h = \sqrt{(1-a^2)}$ . It is important to mention that when dealing with variances and parameters restricted to a finite range such as the autoregressive coefficients is necessary to transform them with the logarithm in the first case or the logit transformation in the latter to use a normal approximation implied in C.2.

## D. General Algorithm

Finally, plugging together RBPF and PKS these are the steps of the general algorithm:

1. Initialisation step to get $D$ draws from:
   a) the parameters $\theta_0^{(j)} \sim p(\theta_0)$ for $j = 1, 2, .., D$.
   b) the states $\left\{f_{0|0}^{(j)}\right\}_{j=1}^{D}$ and their variances $\left\{P_{0|0}^{(j)}\right\}_{j=1}^{D}$, and the volatilities of the factor innovations $\left\{Q_0^{(j)}\right\}_{j=1}^{D}$.
   c) the importance weights $w_0^{(j)} = 1/D$ for $j = 1, 2, ..., D$.
2. For t=1,2,..,T

a) Compute the mean $m_{t-1}^{(j)}$ using C.4 and the variance $V_{t-1}$ of the draws $\theta_{t-1}^{(j)}$.

b) PREDICTION step of the Kalman filter to get the predicted states $\left\{f_{t|t-1}^{(j)}\right\}_{j=1}^{D}$ and their variances $\left\{P_{t|t-1}^{(j)}\right\}_{j=1}^{D}$. We also compute the forecast errors $\left\{v_t^{(j)}\right\}_{j=1}^{D}$ and their variances $\left\{\Sigma_t^{(j)}\right\}_{j=1}^{D}$, and the likelihood $\left\{lik_t^{(j)}\right\}_{j=1}^{D}$ given the observed $y_t$.

c) Compute the IMPORTANCE WEIGHTS for each particle from the likelihood function $\left\{\tilde{w}_t^{(j)}\right\}_{j=1}^{D}$ and normalize them $w_t^{(j)} = \tilde{w}_t^{(j)} / \sum_{j=1}^{D} \tilde{w}_t^{(j)}$.

d) RESAMPLING of the particles using draws from a multinomial $k$ that can take values $1, 2, ..., D$ with probabilities $w_t^{(j)}$, and selecting these particles for the states $\left\{f_{t-1|t-1}^{(k_j)}\right\}_{j=1}^{D}$ and their variances $\left\{P_{t-1|t-1}^{(k_j)}\right\}_{j=1}^{D}$, and the volatilities of the factor innovation $\left\{Q_{t-1}^{(k_j)}\right\}_{j=1}^{D}$.

e) UPDATING step for:

   i. the parameters $\theta_t^{(j)} \sim N(m_{t-1}^{(k_j)}, h^2 V_{t-1})$.

   ii. the state vector $\left\{f_{t|t}^{(j)}\right\}_{j=1}^{D}$ and its variance $\left\{P_{t|t}^{(j)}\right\}_{j=1}^{D}$ using the Kalman filter and the selected particles from the previous step.

   iii. the volatilities of the factor innovations $\left\{Q_t^{(j)}\right\}_{j=1}^{D}$ considering that $\ln Q_t^{(j)} \sim N(\ln Q_{t-1}^{(j)}, \sigma_W^2)$.

   iv. the importance weights $w_t^{(j)} = 1/D$ for $j = 1, 2, ..., D$.