# Business Cycle and Stock Market Volatility: A Particle Filter Approach

Roberto Casarin[†‡¶] and Carmine Trecroci[‡§]

[†]CEREMADE and Dept. of Mathematics, University Paris Dauphine
[‡]Dept. of Economics, University of Brescia
[§]Dept. of Economics, University of Glasgow

First Version: February, 18, 2006

**Abstract**

The recent observed decline of business cycle variability suggests that broad macroeconomic risk may have fallen as well. This may in turn have some impact on equity risk premia. We investigate the latent structures in the volatilities of the business cycle and stock market valuations by estimating a Markov switching stochastic volatility model. We propose a sequential Monte Carlo technique for the Bayesian inference on both the unknown parameters and the latent variables of the hidden Markov model. Sequential importance sampling is used for filtering the latent variables and kernel estimator with a multiple-bandwidth is employed to reconstruct the parameter posterior distribution. We find that the switch to lower variability has occurred in both business cycle and stock market variables along similar patterns.

KEYWORDS: Markov Switching, Stochastic Volatility, Business Cycle, Equity Market, Particle Filters, Sequential Monte Carlo.

JEL CLASSIFICATION: C11, C15, C22, C63, G10, E32, E44.

AMS CLASSIFICATION: 62G07, 62M20, 62P20, 91B84.

[¶]Address: CEREMADE, University Paris IX (Dauphine), 16 place du Maréchal de Lattre de Tassigny, 75775 Paris Cédex 16. E-mail: casarin@ceremade.dauphine.fr. We are grateful to Christian P. Robert for his useful suggestions and to the University of Brescia for financial support. We thank R. Shiller for providing his valuable dataset and participants to the *Computational Statistics and Data Analysis Meeting*, Limassol, Cyprus, 28-31 October 2005 and to *BSMF* seminars for helpful comments.

# 1 Introduction

Influential contributions on the business cycle argue that the early 1980s were a period of structural shifts in both policy conduct (see Clarida, Galì and Gertler (1998)) and business cycle volatility (McConnell and Perez-Quiros (2000); Stock and Watson (2002)). In the US, most statistical accounts of the business cycle identify a marked decline of its volatility since the end of the slowdown of 1980-82. Stock and Watson (2002) dubbed this reduction of volatility as the "Great Moderation". Since 1984-86[1], expansions have lasted longer and slowdowns have been less frequent and shallower than in the previous several decades. On the other hand, the recent swings of aggregate asset prices have set off a wide array of theoretical and empirical views (for instance, Gilchrist, Himmelberg, and Huberman (2005); Menzly, Santos and Veronesi (2004), Shiller (2000)).

The two events/phenomena might be linked. The run-up of stock prices in the late '90s might be related to a reduction of the equity premium. Such fall could be due to a significant decrease of broad macroeconomic risk, as represented by an exogenous decline in business cycle volatility. Indeed, some authors (Lettau et al. (2005)) try to offer explicit rationalizations of this linkage. Now, leaving on a side possible explanations for the widely noticed reduction in the volatility of economic activity (see for example Gordon (2005), Dynan et al. (2006) there is a widespread perception that the relationship between business cycle volatility and financial markets fluctuations is far from straightforward to measure.

There are at least two intertwined reasons why asset prices might experience a boom. A permanent rise in total factor productivity could translate into a persistently higher level of earnings, which in turn could raise, for example, stock market valuations. On the other hand, non-fundamental shocks in the equity or housing markets, perhaps due to over-optimistic expectations about future productivity and returns, could boost prices in the short to medium term. In practice, stock prices represent a measure of the marginal value of firms' installed capital. Greater uncertainty about earnings prospects should immediately translate into more volatile stock prices, while equity and investment goods prices (returns) should therefore both be pro-cyclical (countercyclical) across the economy.

This work tries to assess whether the historical volatility structures of business cycle and stock market fluctuations share some common pattern. We do so by analysing quarterly time series of U.S. stock indices and key macroeconomic variables over the past 40 years, and studying their filtered volatilities. Inter alia, such exercise helps to evaluate the hypothesis that the documented decrease in business cycle volatility has triggered a fall in the equity premium, and then a run-up in stock valuations, at the end of 1990s.

[1]Most studies date 1984-86 as the period in which the decline of volatility took place.

Our choice of a relatively unusual frequency and time span for financial data is motivated by the need to address the issue from a long-term perspective.

In applied statistics (see Perpiñán (2001)) and in macroeconometric studies the use of latent variable models is widely accepted. For instance, one could test the hypothesis that the decline in macroeconomic volatility could affect the investors' perception of macroeconomic risk and the size of the equity premium by fitting a hidden Markov-switching model. This way one could pick up abrupt changes in volatility patterns, and identify common patterns across output (or its components) and stock prices volatility. In this paper we perform a similar exercise to extract latent components in the time series of our interest, by applying the particle filter methodology. More specifically, we follow a Bayesian approach to time series modelling (see Harrison and West (1997), Bauwens, Lubrano and Richard (1999), Kim and Nelson (1999)), and apply sequential Monte Carlo techniques (particle filters) to the joint estimation of the parameters and latent factors.

Particle filters are now widely employed in the estimation of models for financial markets, in particular for stochastic volatility models (see for example Pitt and Shephard (1999) and Lopes and Marigno (2001)). Recently, some studies have proposed their application to macroeconometrics, for the analysis of general equilibrium models (Villaverde and Ramirez (2004a) and (2004b)), the extraction of information from the yield curve (Chopin and Pelgrin (2004)), and for the estimation of latent factors in business cycle analysis (Billio, Casarin and Sartore (2004)). The main advantage of these simulation-based techniques lies in their great flexibility when treating nonlinear dynamic models, which cannot be successfully handled through the traditional Kalman-Bucy or Hamilton-Kitagawa filters. Another advantage comes from the sequential nature of the particle filters, which allows dealing with large datasets and to build on-line applications, in contrast to other simulation-based filtering techniques like Markov Chain Monte Carlo. In this work we follow the sequential Monte Carlo approach proposed by Liu and Chen (1998), and improve upon it by computing a multiple-bandwidth kernel estimate of the parameter posterior distribution.

We detect a parallel decline of macroeconomic and stock market variability, with the switch to low volatility occurring first in the business cycle indicators. More in detail, the volatility of our business cycle indicators follows a low-volatility regime for the second part of the 1966-2003 sample. On the other hand, our estimates identify a similar pattern for the volatility of all our stock market indicators: the market index and its price-earnings and dividend-price ratios all switch to low-volatility around 1991. This switch appears to be persistent, as brief reversions to high volatility occurred only after the longest period of low volatility in our sample. Therefore, we cannot reject the view that the widely observed decline in US business cycle volatility has prompted a similar long-term decrease in stock market volatility.

The paper is structured as follows. Section 2 describes some stylized facts

about the data we use and introduces the simple Bayesian dynamic model employed to identify the latent switching structures for stochastic volatility. Section 3 introduces particle filters for the estimation of the latent factors and discusses some parameter estimation issues. Section 3.2 discusses the multiple bandwidth regularised filter and the convergence of the resulting algorithm. Section 4 comments on our estimation results on synthetic and real data. Section 5 concludes.

# 2 The Volatilities of Economic Activity and Stock Market

## 2.1 Some Stylized Facts

Figure 1 shows the quarterly growth rate of US real personal consumption expenditure per capita (PCE, top chart), and the change in real residential and non-residential fixed investment (RI and NRI, respectively, middle chart), from 1966Q2 to 2003Q3[2]. In all our Figures, grey vertical bars denote NBER-dated contraction episodes. In the case of consumption one can notice an apparent fall of volatility starting from mid-1980s. Aggregate investment data somehow yield the same visual impression. However, before drawing any conclusion a number of factors and events should be accounted for. For instance, the available data span only a limited number of full economic cycles. This is a problem, since they show pronounced swings exactly in correspondence of output fluctuations. The easiest thing to notice is the marked procyclicality of volatilities: inflation outbursts, oil price shocks and known phases of macroeconomic expansions and contractions can all be clearly identified.

Our simple graphical evidence and, more formally, the data analyzed by Gordon (2005), show that the volatilities of residential investment and personal consumption expenditure do follow a clear downward trend. However, in the case of NRI, which is perhaps the most interesting of our investment series, there seems to be no significant level shift. In fact its volatility appears to have been, if anything, slightly higher in the recent past. The absence of significant exogenous shocks and the generalized fall of inflation levels in the eighties and nineties might well account for the observed decrease in PCE and RI volatility (see Gordon (2005)). Also, sizeable portfolio reallocations triggered by the expansion of global liquidity and credit around and after the millennium certainly had some impact on investment choices, as claimed, inter alia, by Borio and Lowe (2004). All in all, however, these data confirm Stock and Watson (2002) conclusions that the cyclical volatility of economic activity in the US has markedly dropped after 1984's.

Against the background of the observed decline in business cycle variability, the perception of macroeconomic risk, as summarized by GDP data, might

[2]The definition and sources of all data are detailed in the Data Appendix.

have decreased since early 1980s as well[3]. Let us therefore look at some stock market valuation data. Figure 1, bottom chart, plots the quarterly real S&P500 price/earnings ratio, defined over a 10-year moving average of earnings, from 1946 to 2004. The increase of equity valuations over the past two decades is apparent, as it is their partial reversion at the end of the nineties. The price-dividend ratio displays a similarly trended pattern, with unprecedented low levels of the dividend yield at the turn of the millennium. Do these data tell us anything about the link between macroeconomic volatility and stock market valuations? Figure 1 also shows the output gap, computed using the potential GDP measure of the Congressional Budget Office (see the Data Appendix). A cursory look at the two series confirms an impressive common behaviour. However, we cannot accept this as evidence of a systematic relationship between their second moments, as the series' mean-reverting properties clearly differ.

Referring to first principles, the classical consumption-based asset pricing model[4] states that risk premia are proportional to the covariance of returns with consumption growth. Let us suppose that the preferences of the representative investor are time-separable:

$$\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}\right) \tag{1}$$

Optimal allocation of resources to consumption and investment implies that the marginal value of wealth and the marginal utility of consumption are equal. The first-order condition of this optimization problem therefore yields

$$\mathbb{E}_t\left(R_{t+1}^{ei}\right) = -\mathbb{C}ov_t\left(R_{t+1}^{ei}, \frac{u'\left(c_{t+1}\right)}{u'\left(c_t\right)}\right), \forall\, i \tag{2}$$

where $R_t^{ei} = R_t^i - R_t^0$ is the return of asset $i$ in excess of a reference asset return $R_t^0$, and $\mathbb{C}ov_t\left(\cdot,\cdot\right)$ denotes conditional covariance.

Using the conventional power utility function, $u\left(c\right) = \frac{c^{1-\gamma}-1}{1-\gamma}$, we can rewrite equation (2) as

$$\mathbb{E}_t\left(R_{t+1}^{ei}\right) = \gamma\mathbb{C}ov_t\left(R_{t+1}^{ei}, \frac{c_{t+1}}{c_t}\right) \tag{3}$$

The last expression states that assets generate expected excess returns for their systematic risk, as summarized by the covariance of returns with the growth in the marginal value of wealth (the so-called *stochastic discount factor*, *pricing kernel*, or *state-price density*). The covariance represents systematic risk because it measures the component of return that contributes to fluctuations in the marginal utility of wealth.

[3]Lettau et al. (2005) point to the same conclusion.

[4]Ferson (2003) and Cochrane (2005) are excellent surveys.

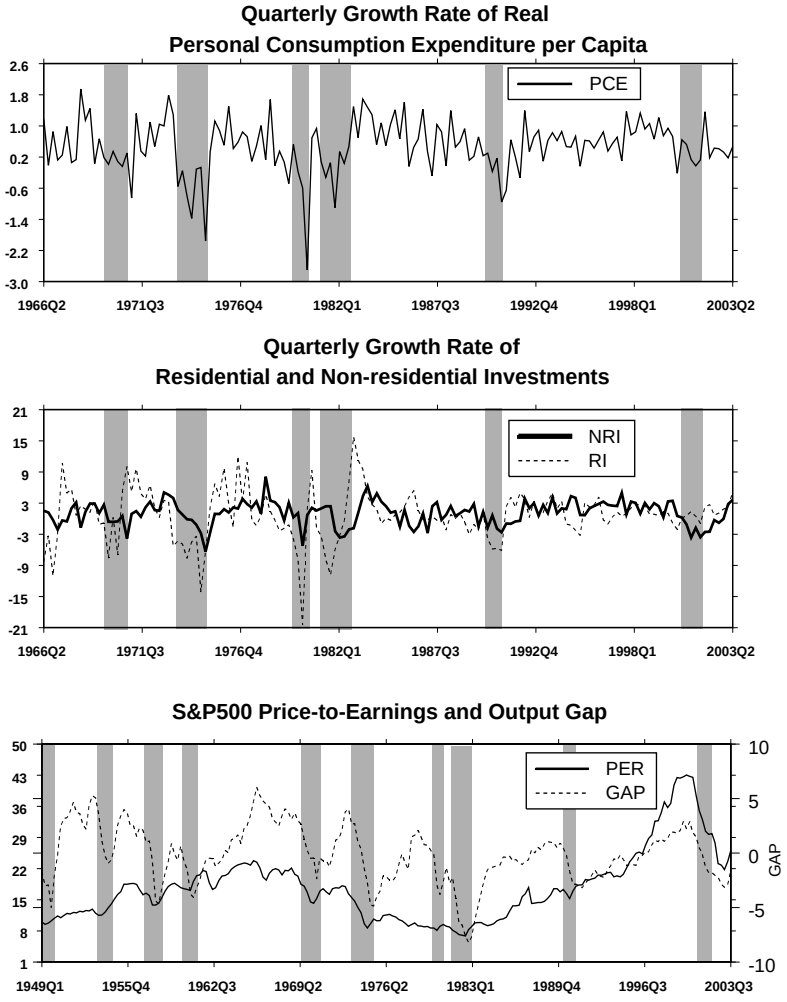


Figure 1: Quarterly growth rate of real personal consumption expenditure per capita (top chart); quarterly growth rate of residential and non-residential investment expenditure (middle chart); S&P500-based price-to-earnings ratio (left-hand scale) and output gap (right-hand scale) (bottom chart). NBER recessions are denoted by shaded bars.

By re-arranging (3) we can write

$$\mathbb{E}_t\left(R_{t+1}^{ei}\right) = \gamma \mathbb{V}_t\left(R_{t+1}^{ei}\right) \mathbb{V}_t\left(\Delta c_{t+1}\right) \rho_t\left(\Delta c, R_{t+1}^{ei}\right) \tag{4}$$

Risk premia therefore vary over time. They move if the conditional variance of excess returns, $\mathbb{V}_t\left(R_{t+1}^{ei}\right)$, changes over time. But fluctuations in risk aversion ($\gamma$), the conditional correlation $\rho_t\left(\Delta c, R_{t+1}^{ei}\right)$, and consumption risk,

$\mathbb{V}_t(\Delta c_{t+1})$ as well explain changes in risk premia[5]. The applied literature (see Cochrane (2005) for a survey) broadly finds that changes in risk and risk aversion at business-cycle frequencies produce predictable excess returns. Therefore, it is clear that investigating the relationships between asset prices and business cycle volatility implies studying the classical issue of the predictability of asset returns and, in essence, the behaviour of risk premia.

In practice, and focusing on equities, aggregate market valuations respond to permanent and transitory changes in expected earnings, in turn affected by the current state of the economy. Expectations about future productivity and returns determine equilibrium on the capital markets: share prices represent a measure of the marginal value of firms' installed capital. On the other hand, changes in share and house prices affect consumption via wealth effects, and alter the rental cost of capital for firms, through a variety of channels. Investment and consumption volatilities might therefore depend on past and expected changes in stock market valuations.

Stock market valuations and aggregate consumption should be cointegrated: rising stock market wealth would lead to rising consumption levels, and vice versa[6]. In fact, Lettau and Ludvigson (2001, 2004) find that the ratio of consumption to wealth forecasts future stock market returns. Moreover, while price/dividend and price/earnings ratios do forecast future share prices returns, they are not leading indicators of consumption growth. Thus, most of short-run stock prices volatility is transitory, which makes the transmission of wealth effects from equity prices to consumption quite a complex process to pin down.

All this suggests that an unrestricted approach, like a univariate approach, to the identification of volatility patterns of stock market indices and business cycle variables could lead to vital information. Moreover, this also motivates the use of low-frequency (quarterly) data.

## 2.2 A Bayesian Dynamic Model

Our broad aim is therefore to check whether common regime changes affect the stochastic volatility profile of some financial and macroeconomic variables. To this goal we adopt latent-variable modelling framework, which is now widely used in statistics and econometrics (see for example Perpiñán (2001) for a review). In structural time series modelling (see Harvey (1989)), the evolution

[5]Note that equation (4) could well be read backwards, with risk premia determing consumption volatility, conditional correlation, etc..

[6]One can show that the class of models we have just sketched provides the fundamental asset pricing equation:

$$P_t = \mathbb{E}_t\left[m_{t+1}\left(P_{t+1} + D_{t+1}\right)\right]$$

where $m$ is the stochastic discount factor ($\equiv \beta\frac{u'(c_{t+1})}{u'(c_t)}$), $P$ is the asset price and $D$ is the amount of any dividends, interest or other payments received by the asset.

of the observed variable is described by means of a set of unobservable (latent) variables that allow capturing time-heterogenous behaviour of the observed series. In the following we consider a Bayesian probabilistic representation of a dynamic model, which is general enough to account for many kinds of latent-variable models.

The proposed probability model consists of the observed process $\{y_t, t \in \mathbb{N}_0\}$, with values in the measurable *observation space* $(E, \mathcal{E})$, $E \subset \mathbb{R}$, and of two latent processes: the volatility regime $\{s_t, t \in \mathbb{N}\}$ and the log-volatility $\{x_t, t \in \mathbb{N}\}$, with values in the measurable *latent product space* $(F \times G, \mathcal{F} \otimes \mathcal{G})$, where $F \subset \mathbb{N}_0$ and $G \subset \mathbb{R}$. Let us introduce $x_{s:t} \stackrel{\Delta}{=} (x_s, \ldots, x_t)$, with $s \leq t$. In the following our Bayesian dynamic model is fully specified through the time-conditional and initial densities.

The latent process $\{s_t, t \in \mathbb{N}\}$, which drives stochastic volatility, is a time-homogeneous Markov chain with transition probabilities

$$p(s_t|s_{0:t-1}, \theta) \sim \mathbb{P}(s_t = k|s_{t-1} = l) = p_{lk} \quad \text{with} \quad k, l \leq K \tag{5}$$

where $K$ is the number of states (or regimes).

Our approach posits that two volatility regimes characterize the behaviour of financial and business cycle variables: low volatility ($s_t = 1$) and high volatility ($s_t = 2$). This is a reasonable assumption, which has been already discussed in many empirical studies on the business cycle (see for example Watson (1994)) and financial variables (see for example So, Lam and Li (1998)). Despite our simple assumption, the model and the proposed inference framework can be easily extended to account for a time-varying number of hidden states, as in Chopin (2001) and Chopin and Pelgrin (2004). Note that, due to the use of a Markov-chain process for the latent variables, the model is also known in the literature as *Hidden Markov Model* (HMM).

The latent log-volatility $\{x_t\}$ and the observation $\{y_t\}$ are parametric processes, with Gaussian conditional densities

$$p(x_t|x_{t-1}, \theta) \sim \mathcal{N}\left(\alpha_{s_t} + \phi x_{t-1}, \sigma^2\right) \tag{6}$$

$$p(y_t|x_t, \theta) \sim \mathcal{N}\left(\mu + \rho y_{t-1}, e^{x_t}\right) \tag{7}$$

where $\theta = (\alpha_1, \alpha_2, p_{11}, p_{22}, \phi, \sigma^2, \mu, \rho)$ is the parameter vector, and $\alpha_{s_t}$ is indexed to the current volatility regime ($s_t = k$). When making Bayesian inference, the parameter vector is treated as a random variable[7]. The initial values of the latent processes are random quantities as well. We assume that the latter follow Bernoulli and Gaussian densities, respectively

$$p(s_0|\theta) \sim \mathcal{B}(1, 0.5) \tag{8}$$

$$p(x_0|\theta) \sim \mathcal{N}\left(\alpha(s_0), \phi\right). \tag{9}$$

[7]Parameter estimation will be discussed in Section 3.1.

Despite its simplicity, the univariate Markov-switching stochastic-volatility (MSSV) model picks up some important stylized facts that characterize financial and macroeconomic time series, such as heteroscedasticity, volatility clustering and switches, and heavy-tails unconditional distributions. The estimation of the MSSV model poses also some challenging statistical problems, as described in the subsequent sections (see also Casarin (2004) for further details).

# 3 Particle Filters

The estimation of the latent-variable model presented in Section 2 configures a problem of nonlinear filtering with unknown parameters. Among the existing methods proposed in the literature, the Bayesian approach reveals to be general enough not only to treat nonlinear problems, but also to account for prior information on the parameters and to allow the use of simulation methods in the inference process.

Let $\mathcal{Y} \subset \mathbb{R}^{n_y}$, $\mathcal{X} \subset \mathbb{R}^{n_x}$ and $\Theta \subset \mathbb{R}^{n_\theta}$ be the observations, state and parameter spaces respectively. We now introduce the following probabilistic representation of a stochastic dynamic model (see Harrison and West (1997))

$$\begin{aligned}
\mathbf{x}_t &\sim p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t-1}, \theta) && \text{(transition density)} && (10)\\
\mathbf{y}_t &\sim p(\mathbf{y}_t|\mathbf{x}_{0:t}, \mathbf{y}_{1:t-1}, \theta) && \text{(measurement density)} && (11)\\
\mathbf{x}_0 &\sim p(\mathbf{x}_0|\theta) && \text{(initial density)} && (12)
\end{aligned}$$

with $t = 1, \ldots, T$, $\mathbf{x}_t \in \mathcal{X}$, $\mathbf{y}_t \in \mathcal{Y}$ and $\theta \in \Theta$. Boldface indicates that state and observation variables could possibly be vectors. For our model $\mathbf{x}_t = (x_t, s_t)$ and $\mathbf{y}_t = y_t$.

By applying the Bayesian theorem and the Chapman-Kolmogorov forward integral equation, we obtain the filtering, one-step-ahead prediction and smoothing densities associated to the dynamic model

$$\begin{aligned}
p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta) &= \frac{p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t-1}, \theta) p(\mathbf{y}_t|\mathbf{x}_{0:t}, \mathbf{y}_{1:t-1}, \theta)}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} && (13)\\
p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}, \theta) &= \int_{\mathcal{X}} p(\mathbf{x}_{t+1}|\mathbf{x}_{0:t}, \mathbf{y}_{1:t}, \theta) p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta) d\mathbf{x}_t && (14)\\
p(\mathbf{x}_s|\mathbf{y}_{1:t}, \theta) &= p(\mathbf{x}_s|\mathbf{y}_{1:s}, \theta) \int_{\mathcal{X}} \frac{p(\mathbf{x}_{s+1}|\mathbf{x}_s, \theta) p(\mathbf{x}_{s+1}|\mathbf{y}_{1:t}, \theta)}{p(\mathbf{x}_{s+1}|\mathbf{y}_{1:t}, \theta)} d\mathbf{x}_{s+1}, && (15)
\end{aligned}$$

with $s < t$. The one-step-ahead prediction density of $\mathbf{y}_t$ is $p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$, and is defined as follows

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \int_{\mathcal{X}} p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t-1}, \theta) p(\mathbf{y}_t|\mathbf{x}_{0:t}, \mathbf{y}_{1:t-1}, \theta) d\mathbf{x}_t. \qquad (16)$$

Note that the analytical solution of the discrete-time filtering problem exists in a few cases: the linear and Gaussian dynamic model and the linear

model with a countable finite-dimension latent space, where Kalman-Bucy (see Kalman (1960)) and Hamilton-Kitagawa (see Hamilton (1989)) filters respectively apply. Except for these cases, the filtering problem is hard to solve, and approximation methods are usually applied for the more difficult nonlinear problems.

The simulation-based Bayesian framework represents an appealing solution to nonlinear estimation problems. Thanks to a probabilistic representation of the model of interest, the Bayesian approach allows including advanced Monte Carlo techniques into the inference process. For example, Harrison and West (1997) propose a Bayesian state-space form that permits to deal with both discrete and continuous-valued state space models. Kim and Nelson (1999) employ a Bayesian representation of the hidden Markov models for business cycle analysis. All these works address the latent factor extraction problem through the use of *Monte Carlo Markov Chain* (MCMC) techniques (see also Robert and Casella (2004)), such as Gibbs sampling, although in Harrison and West (1997) an earlier adaptive importance sampling algorithm is proposed as an alternative to MCMC. In Bayesian analysis MCMC techniques are considered as the most suitable tool for solving integration problems that arise in parameters and latent-variables estimation, and in hypothesis testing. Nevertheless, many applications reveal that MCMC methods have some drawbacks too. For example, the choice of the scale parameter of the random walk proposal in a Metropolis-Hastings algorithm can severely affect the convergence to the posterior distribution. Moreover, the generated Markov chain can get trapped in a local mode of the posterior distribution.

To estimate the hidden states and the parameters of our hidden Markov model, we follow a route alternative to MCMC, which relies upon sequential Monte Carlo methods. This class of methods is referred in the literature as to *particle filters* (also known as *bootstrap filters*, *interacting particle filters*, *condensation algorithms*, *Monte Carlo filters*). It was first introduced for nonlinear filtering simultaneously by Gordon, Salmond and Smith (1993), and Berzuini et al. (1997) (see also Doucet, Freitas and Gordon (2001) for a review of sequential Monte Carlo methods). Particles filters are simulation-based devices and reveal to be quite useful for filtering in complex dynamic models. They make use of a weighted Monte Carlo sample to cover the whole state space and to easily update, over the time iterations, the parameter posterior distribution. The sequential nature of the filter also allows capturing or detecting sudden changes of the states and parameters of the system, which is a feature particularly helpful in our context.

We now present the baseline particle filter method, also called *Sequential Importance Sampling* (SIS). Assume that a weighted Monte Carlo sample (also called *particle set*), $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^N$, from the posterior $p(\mathbf{x}_t|\mathbf{y}_{1:t}, \theta)$, is available at

time $t$. It represents the empirical prior for the states at time $t+1$, i.e.

$$p_N(\mathbf{x}_t|\mathbf{y}_{1:t},\theta) = \sum_{i=1}^{N} w_t^i \delta_{\{\mathbf{x}_t^i\}}(\mathbf{x}_t), \tag{17}$$

where $\delta_{\{y\}}(x)$ denotes the Dirac point-mass centered in $y$. Using the empirical prior, the one-step-ahead prediction density in Eq. (14) and the filtering density in Eq. (13) can be approximated as follows

$$\begin{aligned} p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t},\theta) &= \int_{\mathcal{X}} p(\mathbf{x}_{t+1}|\mathbf{x}_t,\theta) p(\mathbf{x}_t|\mathbf{y}_{1:t},\theta) d\mathbf{x}_t \\ &\approx \sum_{i=1}^{N} w_t^i p(\mathbf{x}_{t+1}|\mathbf{x}_t^i,\theta) \qquad (18) \\ p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1},\theta) &\propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta) p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t},\theta) \\ &\approx \sum_{i=1}^{N} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta) p(\mathbf{x}_{t+1}|\mathbf{x}_t^i,\theta) w_t^i \qquad (19) \end{aligned}$$

where the first summation defines the *empirical prediction density*, $p_N(\mathbf{x}_{t+1}|\mathbf{y}_{1:t},\theta)$, and the second one the *empirical filtering density*, $p_N(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1},\theta)$.

A simple way to obtain a weighted sample, $\{\mathbf{x}_{t+1}^i, w_{t+1}^i\}_{i=1}^N$, from the filtering density at time $t+1$, is to apply importance sampling to the empirical filtering density (Eq. (19)). A natural choice for the importance density is the transition density, because it represents a sort of prior at time $t$ for the state $\mathbf{x}_{t+1}$. The particle weights are updated as follows

$$\begin{aligned} \tilde{w}_{t+1}^i &\propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta) p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t},\theta) w_t^i}{q(\mathbf{x}_{t+1}|\mathbf{x}_t^i,\mathbf{y}_{t+1},\theta)} = w_t^i \, p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^i,\theta) \qquad (20) \\ w_{t+1}^i &= \frac{\tilde{w}_{t+1}^i}{\sum_{j=1}^N \tilde{w}_{t+1}^j} \qquad (21) \end{aligned}$$

It is well known (see for example Arulampalam et al. (2001)) that after some iterations of the SIS algorithm, the empirical distribution degenerates into a single particle, as the variance of the importance weights is non-decreasing over time (see Doucet (2000)). In order to solve the degeneracy problem, many extensions to the SIS have been proposed in the literature: the *Sampling Importance Resampling* in Gordon, Salmond and Smith (1993), the *Regularised Particle Filter* in Musso, Oudjane and LeGland (2001) and the *SIR-Move* algorithm in Gilks and Berzuini (2001).

In this work we follow Pitt and Shephard (1999), who introduce the *Auxiliary Particle Filter* (APF). In order to avoid a pure re-sampling step, the APF algorithm uses an auxiliary variable to select the most representative particles and to mutate them through a simulation step. The weights of

the regenerated particles are updated through importance sampling. In this way particles with low probability do not survive to the selection, and the information contained in the particle set is not wasted. In particular, the auxiliary variable is a random particle index, which is used in the resampling step to select the new particles. The random index is simulated from a discrete distribution which reflects the information based on the previous particle set and the new observation $y_{t+1}$. Note that the empirical filtering density given in Eq. (19) is a mixture of distributions, which can be demarginalized as follows

$$\begin{aligned} p(\mathbf{x}_{t+1}, i|\mathbf{y}_{1:t+1}, \theta) &= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \theta)} p(\mathbf{x}_{t+1}|i, \mathbf{y}_{1:t}, \theta) p(i|\mathbf{y}_{1:t}, \theta) \quad (22) \\ &= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}, \theta)} p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta) w_t^i. \end{aligned}$$

where $i \in \{1, \ldots, N\}$ is an allocation variable, which selects one of the mixture components. The algorithm generates a new set of particles by jointly simulating the particle index $i$ (*selection step*) and the selected particle value $\mathbf{x}_{t+1}$ (*mutation step*) from the reparameterised empirical filtering density, and according to the following importance density

$$q(\mathbf{x}_{t+1}^j, i^j|\mathbf{y}_{1:t+1}, \theta) = q(\mathbf{x}_{t+1}^j|\mathbf{x}_t^{i^j}, \theta) q(i^j|\mathbf{y}_{1:t+1}, \theta)$$

for $j = 1, \ldots, N$ and with $q(i^j|\mathbf{y}_{1:t+1}, \theta) = p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, \theta) w_t^{i^j}$. By following the usual importance sampling argument, the updating relation for the particle weights is

$$w_{t+1}^j \triangleq \frac{p(\mathbf{x}_{t+1}^j, i^j|\mathbf{y}_{1:t+1}, \theta)}{q(\mathbf{x}_{t+1}^j, i^j|\mathbf{y}_{1:t+1}, \theta)} = \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j, \theta)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j}, \theta)}.$$

In many applications of the particle filter techniques, parameters are treated as known, and MCMC parameter estimates are used in place of the true values. MCMC is typically an off-line approach, as it does not allow to sequentially update parameter estimates as new observations arrive. Moreover, when applied sequentially, MCMC estimation is more time consuming than particle filter algorithms. Thus in the next section we will consider the filtering problem in presence of unknown static parameters, in a Bayesian perspective.

### 3.1 Sequential Parameter Estimation

In economic applications both the latent variables and the parameters are unknown quantities. Therefore, both the problems of hidden state filtering and parameter estimation arise. In this work we propose a full Bayesian sequential estimation approach for the parameter of interest. Parameters can

be estimated jointly or separately w.r.t. the state filtering, as evidenced by

$$p(\mathbf{x}_{t+1}, \theta|\mathbf{y}_{1:t+1}) = \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t, \theta)p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})}p(\theta|\mathbf{y}_{1:t}) \quad (23)$$

$$= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \mathbf{y}_{1:t}, \theta)p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}_{1:t}, \theta)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})}p(\mathbf{x}_{0:t}, \theta|\mathbf{y}_{1:t}). \quad (24)$$

Equation (23) shows that the filtering problem can be treated conditionally on the parameters. It is possible for example to use the Kalman Filter or the HMM filtering algorithms to estimate the states and the particle filter to estimate the parameters (see for example Chopin (2001)). In our MSSV model neither the Kalman filter nor the HMM can be used, thus Monte Carlo filters could be applied for the joint estimation of parameters and states of the dynamic system.

Equation (24) shows the recursive relation for the joint state and parameter filtering problem. It also suggests a direct way to estimate parameters, by including them in the state vector. Berzuini et al. (1997) develop this approach in a Bayesian framework and apply standard particle filtering techniques to estimate the joint posterior density $p(\mathbf{x}_{0:t}, \theta|\mathbf{y}_{1:t})$. The approximated posterior $p(\theta|\mathbf{y}_{1:t})$ is then obtained by marginalisation. This approximation requires that the whole history of the particle set is stored and has a computational complexity of $\mathcal{O}(Nt)$. Rather, when available, a set of sufficient statistics for the likelihood function can be alternatively updated and stored to approximate the posterior.

In the naïve method proposed in Berzuini et al. (1997), the static parameters are included in the state vector. As pointed out by Storvik (2002), the absence of dynamics in the parameters produces a negative effect on the estimation of the posterior distribution, which degenerates into a Dirac mass. Different solutions to the degeneracy problem have been proposed in the literature[8].

In the following we refer to the framework due to Liu and West (2001), who apply a regularisation technique similar to the one used by Musso, Oudjane and LeGland (2001). In our algorithm the posterior distribution $p(\theta_t|\mathbf{y}_{1:t})$ is approximated by means of the particle set $\{\mathbf{x}_t^i, \theta_t^i, w_t^i\}$, where the index $t$ of the parameter vector means that it is updated sequentially. The resulting empirical distribution

$$\begin{aligned} &p_N(\mathbf{x}_{t+1}, \theta_{t+1}|\mathbf{y}_{1:t+1}) = \\ &= \sum_{i=1}^{N} \int_{\mathcal{X}\times\Theta} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta_{t+1})p(\mathbf{x}_{t+1}|\mathbf{x}_t, \theta_{t+1})\delta_{\{\theta_t\}}(\theta_{t+1})w_t^i\delta_{\{(\mathbf{x}_t^i, \theta_t^i)\}}d\mathbf{x}_t d\theta_t \\ &= \sum_{i=1}^{N} p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \theta_t^i)p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \theta_t^i)\delta_{\{\theta_t^i\}}(\theta_{t+1})w_t^i \end{aligned} \quad (25)$$

[8] See for example Kitagawa (1998), Storvik (2002) and Polson, Stroud and Müller (2002, 2003).

is then regularised through a multiple-bandwidth Gaussian-kernel density estimation

$$\frac{1}{N}\sum_{i=1}^{N} w_t^i p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta_t^i)p(\mathbf{x}_{t+1}|\mathbf{x}_t^i,\theta_t^i)\mathcal{N}_{n_\theta}(\theta_{t+1}|m_t^i,(b_t^i)^2 V_t) \tag{26}$$

where $(b_t^i)^2 = (1-(a_t^i)^2)$ and $m_t^i = \theta_t^i a_t^i + \bar{\theta}(1-a_t^i)$. Due to the kernel reconstruction (also called *regularisation step*), the original parameter transition density, $\delta_{\{\theta_t^i\}}(\theta_{t+1})$, has been replaced by a Gaussian transition density $\mathcal{N}_{n_\theta}(\theta_{t+1}|m_t^i,(b_t^i)^2\,V_t)$ with particle-specifics: scale factor, $b_t^i$, location, $m_t^i$ and shrinkage factor, $a_t^i$. The choice of $b_i^t$ and $a_i^t$ will be discussed in section 3.2.

The reconstruction of the posterior distribution through Gaussian kernel density estimation is a technique introduced by West (1992, 1993) in order to obtain an adaptive importance sampling algorithm. The use of adapting importance functions is particulary suitable in the estimation of nonlinear dynamic models, where the probability density function of the system is not time-homogeneous.

After the kernel reconstruction of the posterior density, a new set of particles can be generated by applying the APF algorithm to the states and the parameters using the posterior density kernel estimate as importance density.

In the spirit of the APF algorithms, the empirical posterior distribution is reparameterised, using the auxiliary variable $i$ to select the mixture component

$$\hat{p}_N(\mathbf{x}_{t+1},\theta_{t+1},i) = w_t^i p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1},\theta_t^i)p(\mathbf{x}_{t+1}|\mathbf{x}_t^i,\theta_t^i)\mathcal{N}_{n_\theta}(\theta_{t+1}|m_t^i,(b_t^i)^2\,V_t) \tag{27}$$

where $\hat{p}_N(\mathbf{x}_{t+1},\theta_{t+1}|\mathbf{y}_{1:t+1})$ denotes the multiple-bandwidth kernel estimator. The particle selection step is obtained by sampling the mixture index $i$ together with states $\mathbf{x}_{t+1}$ and parameters $\theta_{t+1}$. A sample from the joint density (27) is obtained through importance sampling with importance density

$$q(\mathbf{x}_{t+1},\theta_{t+1},i|\mathbf{y}_{1:t+1}) = p(\mathbf{x}_{t+1}|\mathbf{x}_t^i,\theta_t^i)\mathcal{N}_{n_\theta}(\theta_{t+1}|m_t^i,(b_t^i)^2\,V_t)q(i|\mathbf{y}_{1:t+1}) \tag{28}$$

where the importance function used to sample the random index is $q(i|\mathbf{y}_{1:t+1}) = p(\mathbf{y}_{t+1}|\mu_{t+1}^i,m_t^i)w_t^i$. From previous assumptions the weights update as follows

$$\begin{aligned} w_{t+1}^j &\propto \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j,\theta_t^j)p(\mathbf{x}_{t+1}^j|\mathbf{x}_t^{i^j},\theta_t^j)\mathcal{N}_{n_\theta}(\theta_{t+1}^j|m_t^{i^j},(b_t^i)^2\,V_t)w_t^{i^j}}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j},m_t^{i^j})p(\mathbf{x}_{t+1}^j|\mathbf{x}_t^{i^j},\theta_t^{i^j})\mathcal{N}_{n_\theta}(\theta_{t+1}^j|m_t^{i^j},(b_t^i)^2\,V_t)w_t^{i^j}} \\ &= \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}^j,\theta_t^j)}{p(\mathbf{y}_{t+1}|\mu_{t+1}^{i^j},m_t^{i^j})}, \end{aligned} \tag{29}$$

with $j=1,\ldots,N$.

### 3.2 Multiple-Bandwidths Regularisation and Convergence Issues

The particle filter given in the previous section avoids degeneracy in parameter estimates by introducing variability in the particle set through a regularisation step. We stress again that, in Eq. (26), for each particle the original transition equation, $\theta_{t+1} = \theta_t^i$ a.s., is replaced by a $n_\theta$-dimensional Gaussian autoregressive process

$$\theta_{t+1} = a_t^i \theta_t^i + (1 - a_t^i)\bar{\theta}_t + b_t^i V_t^{1/2} \xi_{t+1}, \text{ with } \xi_{t+1} \sim \mathcal{N}_{n_\theta}(0, Id),$$

where $a_t^i$ is a shrinkage factor, $b_t^i$ a local scale factor and $Id$ is the $(n_\theta \times n_\theta)$ identity matrix.

The shrinkage technique between Gaussian kernels is due to Liu and West (2001) and allows us to reduce the negative effects of the variability introduced with the artificial evolution of the parameter. The kernel density at time $t+1$ is made dependent on the density at time $t$ through the constraint on the conditional variance: $\mathbb{V}_t(\theta_{t+1}) = \mathbb{V}_t(\theta_t) \triangleq V_t$. Under this constraint and assuming that the variance-covariance matrix of the noise is proportional to $V_t$ and to a particle-specific discount factor $\lambda_t^i$ (i.e. $\mathbb{C}ov_t(\xi_{t+1}, \theta_t) = V_t((\lambda_t^i)^{-1} - 1)$), the conditional scale and location of the transition density are

$$\mathbb{E}_t(\theta_{t+1}) = a_t^i \, \theta_t^i + (1 - a_t^i)\mathbb{E}_t(\theta_t), \quad \mathbb{V}_t(\theta_{t+1}) = (b_t^i)^2 V_t, \tag{30}$$

where $a_t^i = (3\lambda_t^i - 1)(2\lambda_t^i)^{-1}$ and $(b_t^i)^2 = (1 - (a_t^i)^2)$. It follows that each component of the kernel density estimator has its own specific location and scale factor. Conditional mean and variance can be replaced by their estimates based on the particles set available at time $t$: $\bar{\theta}_t = N^{-1} \sum_{i=1}^N \theta_t^i$ and $\hat{V}_t = (N(N-1))^{-1} \sum_{i=1}^N (\theta_t^i - \bar{\theta}_t)'(\theta_t^i - \bar{\theta}_t)$.

Two important remarks should be done about the choice of the discount factor $\lambda$. The first remark concerns the relation between the discount factor and the bandwidth $b$. The choice of $\lambda$ should be driven by one of the criteria proposed in the literature for determining the optimal bandwidth $b$. For example, for Gaussian data and single-bandwidth Gaussian kernels the optimal bandwidth in the integrated MSE sense is

$$b_N^* = \left( \frac{4}{(2 + n_\theta)N} \right)^{\frac{1}{n_\theta + 4}}$$

which is also used as a rule of thumb for non-Gaussian data. In our multiple-bandwidth setting, the sequence of bandwidths is in the interval $[(b_N^*)^\alpha, (b_N^*)^\beta]$, with $\alpha < 1$ and $\beta > 1$. By applying to the boundaries the following monotone decreasing relation between the discount factor and the bandwidth,

$$\lambda_N = \frac{3 + 2\sqrt{1 - (b_N)^2}}{4(b_N)^2 + 5}$$

we find two boundaries: $\underline{\lambda}$ and $\bar{\lambda}$, for the discount factor. Note that the two values depend on the dimensions of the particle set and of the parameter space.

As for the second remark, our numerical experiments suggest that the choice of $\lambda$ clearly affects the convergence of the estimates and is itself dependent on the employed number of particles and also on the parameter setting of the dynamic model under study.

In order to account for both the previous instances, in this work, we propose a full Bayesian approach and include $\lambda$ in the parameter vector with a prior distribution on it. The initial set of $N$ particles, $\{\lambda_0^i\}_{i=1}^N$, is simulated from a uniform prior distribution: $\mathcal{U}(\lambda; \underline{\lambda}_N, \bar{\lambda}_N)$. A truncated Gaussian distribution, $\mathcal{TN}(\lambda; \underline{\lambda}_N, \bar{\lambda}_N, m, 1)$, is used as transition density in the mutation step of the particle filter. Although a time-dependent discount factor could be used or a more elaborate self-tuning iterative procedure of the discount factor could be proposed, in this work we simply suggest to keep the advantages of using a multiple-bandwidth estimator in the filtering procedure and focus on some basic convergence properties of the resulting algorithm.

Allowing the discount factor to vary over the particle set (see last line in Eq. (26)), amounts performing a multiple-bandwidth kernel estimation of the posterior density. Thus, the regularisation step introduced in our algorithm belongs to the general class of *sample-point* kernel density estimators. Sample-point estimators have locally adapted bandwidths and show to be more promising than fixed bandwidth estimators. Fixed bandwidths deal badly with local scale variations of the prediction density and produce undersmoothing of the tails and oversmoothing of the peaks.

To prove the convergence of the kernel particle filter with respect to the number of particle $N$, we refer to the following $n_\theta$-dimensional sample-point kernel estimator of $p(\theta_{t+1}|y_{1:t+1})$

$$\hat{p}_N(\theta_{t+1}|y_{1:t+1}) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{(b_t^i)^{n_\theta}} K((b_t^i)^{-n_\theta}(\theta_{t+1} - \theta_t^i)) \tag{31}$$

where $b_t^i$ is a positive smoothing parameter associated to the $i$-th particle, which could vary with the algorithm iterations.

The convergence of the basic particle filter has been studied in Crisan (2001) and Crisan and Doucet (2000), (2002)[9].

Let us focus on the sample-point kernel particle filter without selection step. The $L_2$-convergence of the estimator w.r.t. the number of particles $N$, is given in the following.

**Theorem 3.1.** *(Quadratic-mean convergence)*
*Let $K$ be a Gaussian kernel on $\mathbb{R}^{n_\theta}$, $\{b_t^i\}_{i=1}^N$ a sequence of positive scale factors*

[9]For a deeper theoretical analysis of the particle methods see also Del Moral (2004) and Bartoli and Del Moral (2001). The convergence of the Gaussian regularised particle filter of Liu and West (2001) is proved in Stavropoulos and Titterington (2001).

*(bandwidths) with value in the finite interval* $[b_{N,\min}, b_{N,\max}]$ *and* $\{\theta_t^i\}_{i=1}^N$ *a sequence of* $\mathbb{R}^{n_\theta}$*-valued i.i.d. samples simulated from* $\hat{p}(\theta_t|y_{1:t})$*. If*

$$\lim_{N\to\infty} b_{N,\max} = 0, \quad \textit{and} \quad \lim_{N\to\infty} (b_{N,\min})^d N = \infty \tag{32}$$

*then the functional estimator,* $\hat{p}_N(\theta_{t+1}|y_{1:t+1})$ *defined in (31), converges in quadratic mean to the true density* $\hat{p}(\theta|y_{1:t+1})$

$$\hat{p}_N(\theta_{t+1}|y_{1:t+1}) \xrightarrow[N\to\infty]{L^2} p(\theta_{t+1}|y_{1:t+1}). \tag{33}$$

*Proof.* See Appendix C. □

In Appendix C we also prove the a.c. convergence and the a.s. convergence of the kernel estimator. The two imply the $L_1$-convergence that is given in the following.

**Theorem 3.2.** *($L_1$-convergence)*
*Let* $K$ *be a Gaussian kernel on* $\mathbb{R}^{n_\theta}$*,* $p(\theta_t|y_{1:t}) \in L^1(\mathbb{R}^{n_\theta})$ *a density,* $\{b_t^i\}_{i=1}^N$ *a sequence of positive scale factors (bandwidths) with value in the interval* $[b_{N,\min}, b_{N,\max}]$ *and* $\{\theta_t^i\}_{i=1}^N$ *a sequence of* $\mathbb{R}^{n_\theta}$*-valued i.i.d. samples simulated from* $p(\theta_t|y_{1:t})$*. If*

$$\lim_{N\to\infty} b_{N,\max} = 0\,, \quad \lim_{N\to\infty} \frac{(b_{N,\min})^d N}{\log N} = \infty \tag{34}$$

*then the sample-point estimators,* $p_N(\theta_{t+1}|y_{1:t+1})$ *defined in (31), converges in* $L_1$ *to the true density*

$$\hat{p}_N(\theta_{t+1}|y_{1:t+1}) \xrightarrow[N\to\infty]{L^1} p(\theta_{t+1}|y_{1:t+1}). \tag{35}$$

*Proof.* See Appendix C. □

Our convergence results can also be extended to account for the selection step in the particle filter. See for example Rossi (2004) for an updated analysis of the convergence of the kernel-convolution particle filters.

# 4 Estimation Results

In this section we apply our particle filter to the MSSV model given in Eq. (5), (6) and (7). In order to verify the efficiency of the algorithm in the estimation of parameters and hidden variables, and also to detect possible degeneracy problems, we first apply the regularised APF to synthetic data. The sequential Monte Carlo filter is subsequently applied to actual US stock market and business cycle data.

### 4.1 Simulated Data

In the simulations we refer to the MSSV model given in Section 2 and apply our regularised particle filter with multiple bandwidth. The resulting filtering procedure, given in Appendix A, allows us to sequentially estimate the parameter vector and filter and predict the hidden state vector of the dynamic model. We use a set of $N = 2,000$ particles to obtain the empirical filtering and prediction densities[10].

Figure 2 plots on-line estimates of the latent factors $\{x_t\}$ and $\{s_t\}$ obtained form a typical run of the APF algorithm on the synthetic dataset plotted in the first panel of the same figure. In order to detect the absence of degeneracy in the output of the APF algorithm we evaluate at each time step the *survival rate*. This is defined as the number of particles surviving to the selection step over the total number of particles. Particles sets degenerate when persistently exhibiting a high number of dead particles from a generation to the subsequent one. The survival rate is therefore computed as follows

$$SR_t = 1 - \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{\{J_{i,t}=\emptyset\}} \tag{36}$$

where $J_{i,t} = \{j \in \{1, \ldots, N\} | i_t^j = i\}$ is the set of all random index values, which are selecting, at time $t$, the $i$-th particle. If at time $t$ the particle $k$ does not survive to the selection step, then the set $J_{k,t}$ becomes empty. Figure 2 shows the survival rate at each time step. The rate does not decline, therefore we conclude that the APF algorithm does not degenerate in our simulation study.

Figure 3 shows sequential estimates of parameters $\alpha_1$, $\alpha_2$, $p_{11}$, $p_{22}$, $\phi$ and $\sigma^2$. We remind that these are the parameters of the conditional densities of the latent log-volatility $\{x_t\}$ and the observation $\{y_t\}$[11].

Finally, for each filtering iteration Figure 4 shows the empirical mean of the discount factors $\lambda$ used in the regularisation step. The bounds of the interval spanned by the discount factor are represented in the same graph.

### 4.2 Volatility Patterns in Macroeconomic and Stock Market Data

The aim of this exercise is to capture similarities in the volatility profiles of macroeconomic and stock market variables. Our attention therefore focuses in particular on what happened to the variability of stock market valuations over the past three decades, during which volatility of several macroeconomic variables has clearly declined.

[10] All computations have been carried out on a Pentium IV 2.4 Ghz, and the APF algorithm has been implemented in GAUSS 7.0.

[11] See Section 2.2 above.

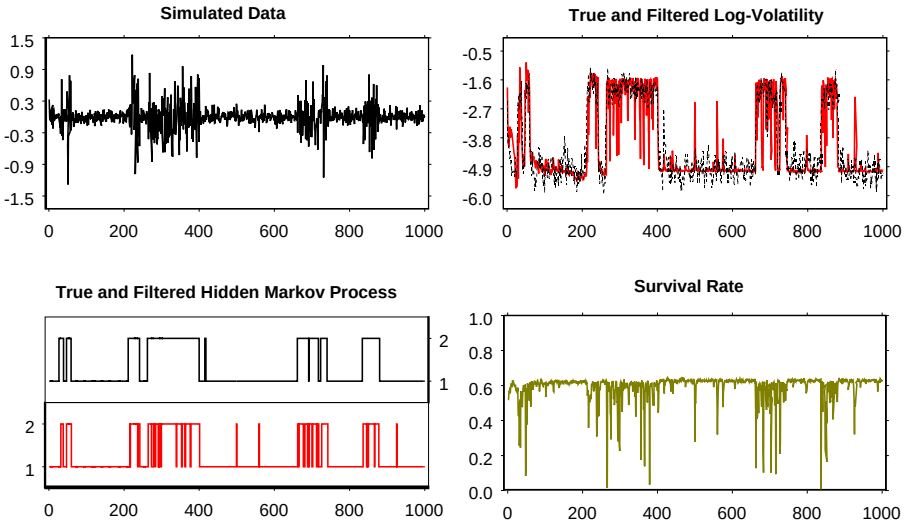


Figure 2: The simulated (black line) and sequentially filtered (grey line) latent factors and the survival rate of the particle set, over a sample of $T = 1,000$ observations.

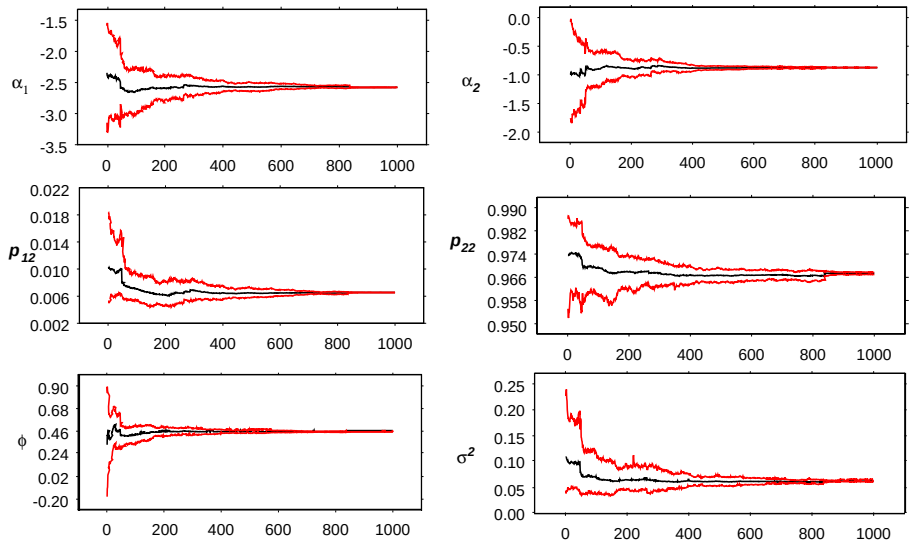


Figure 3: On-line parameter estimates. Graphs show at each date the empirical mean and the quantiles at 0.025 and 0.975 for each parameter.

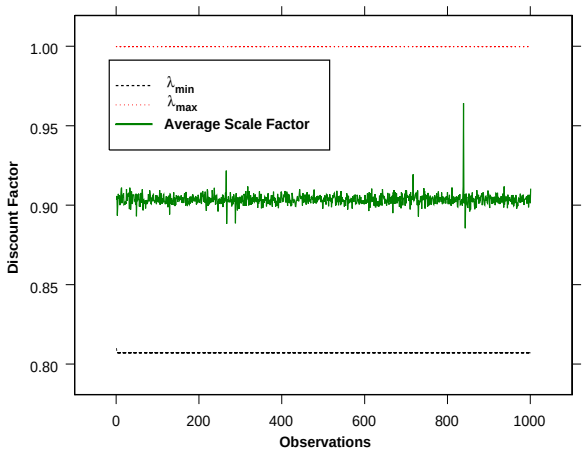


Figure 4: The support set $[\lambda_{\min}, \lambda_{\max}]$ of the random discount factor and the average discount factor over $1,000$ particle filter iterations. We use a set of $N = 2,000$ particles.

We apply the sequential Monte Carlo filter to two groups of actual US data. The first group (the log-difference of the S&P500 stock market index and its corresponding dividend yield and price-to-earnings ratio) tracks the aggregate behaviour of US stock market. The second group of variables (the log-difference of industrial production, non-residential investment expenditure, real personal consumption expenditure per capita, and the level of the output gap) accounts for the evolution of the business cycle[12].

We start by analyzing the behaviour of stock market valuations. We therefore first estimate the MSSV model on the log-difference of the Standard & Poor's 500 index (S&P), and of its corresponding dividend yield (DP) and price-to-earnings ratio (PE). The data are quarterly observations collected on the sample period 1966Q2-2003Q3 (see the Data Appendix for details on data sources and definitions).

The filtered volatility regimes and log-volatility processes are shown in Fig. 5, 6 and 7. The top left-hand panel of each figure plots the time series accompanied by vertical shaded areas representing the contraction phases of the US GDP as detected by Business Cycle Dating Committee of the NBER. The contraction starts at the peak of the cycle and ends at the through. In particular, our sample covers the following GDP contraction periods (the corresponding quarter is in parenthesis) that are of interest in our study: December 1969 (IV)-November 1970 (IV), November 1973 (IV)-March 1975 (I), January 1980 (I)-July 1980 (III), July 1981 (III)-November 1982 (IV), July 1990 (III)-March 1991 (I) and March 2001 (I)-November 2001 (IV).

[12]Gordon (2005) finds that government spending and net exports too might have played a role (albeit a relatively small one) in explaining the decline of output volatility.

From the charts in Figures 5 to 7 one can easily notice that S&P, PE and DP all follow remarkably similar volatility patterns. The estimated filtered log-volatility and filtered Markov processes show that the regimes of high and low volatility alternate along very similar sequences across the three variables. First, S&P, PE and (to a slightly smaller extent) DP all see periods of low volatility that last much longer than those of high volatility. The last twenty years in particular seem to have witnessed only few and short reversions to high-volatility regimes. Second, the vertical bars point to volatility rising almost exclusively around periods of GDP contraction. The 1990-1991 and 2001 GDP recessions triggered a shift to high volatility for S&P, PE and DP, although for the latter the high-volatility regime predates the start of GDP contraction, and for S&P our algorithm detects a similar regime for the latter part of 1990s. Third, volatility switches appear to happen less frequently since the mid-eighties. We could take the latter two findings as *prima facie* evidence of some relationship between lower stock market volatility and decreased GDP volatility.

For all the series the survival rate is well-behaved, therefore pointing to the absence of degeneracy phenomena in the posterior distribution. Figures 13 to 19 in Appendix D show the sequential estimates of the parameter $\alpha_1$, $\alpha_2$, $p_{12}$, $p_{22}$, $\phi$ and $\sigma^2$ with the 0.975 and 0.025 quantiles. The values of the parameter estimates at the last iteration of the filter are in Table 1.

Turning to the examination of business cycle variables, we study the volatility of the following series: Industrial Production (IP), Personal Consumption Expenditure per capita (PCE), Non-Residential Investments (NRI), all in their log-differences, and the level of Output Gap (YGAP)[13]. The filtered stochastic log-volatility and volatility-regime processes are shown in Figures 8, 9, 10 and 11. At the outset, one can notice that the clearest similarity in the behaviour of volatility concerns industrial production and non-residential investment. The volatilities of IP and NRI appear to have switched to a persistent low-volatility regime in the second half of the 1980s. IP reverts to high volatility only briefly after the 1990-91 contraction, and NRI does the same around the 2001 recession and the subsequent recovery. The volatility of PCE too enters a low-level regime, but only since 1993. The fact that in the preceding part of the sample high-volatility dominates more than for the other macroeconomic series makes its switch to low volatility even more striking.

Finally, let us turn to the output gap. We remind that we study its level. Even a bird's eye view of the series (top left-hand chart of Figure 11) confirms that since the mid-eighties expansions have had much longer durations, while slowdowns have both occurred less often and been milder than in the preceding

[13]We do not include here results for residential fixed investment expenditure. Those data, which are notoriously linked to the value of aggregate wealth, will be the object of a multivariate extension.

Table 1: Parameter estimates (posterior-mean estimates) with the 0.975 and 0.025 quantiles, for quarterly log-differences of the observations (except for the Output Gap) on the sample period 1966Q3-2003Q3. Give $h_t$ as the log-volatility, our model $h_t = \alpha_1 \mathbb{I}_{\{1\}}(s_t) + \alpha_2(1 - \mathbb{I}_{\{1\}}(s_t)) + \phi h_{t-1} + \sigma\varepsilon_t$ with $\varepsilon_t$ i.i.d. standard normals.

| | S&P Index | | | Price-to-Earnings Ratio | | |
|---|---|---|---|---|---|---|
| $\theta$ | $\hat{\theta}$ | $q_{0.025}$ | $q_{0.975}$ | $\hat{\theta}$ | $q_{0.025}$ | $q_{0.975}$ |
| $\alpha_1$ | -3.264 | -3.378 | -3.115 | -3.183 | -3.303 | -3.057 |
| $\alpha_2$ | -2.747 | -2.866 | -2.634 | -2.862 | -2.952 | -2.781 |
| $p_{12}$ | 0.014 | 0.012 | 0.017 | 0.018 | 0.017 | 0.020 |
| $p_{22}$ | 0.990 | 0.989 | 0.991 | 0.974 | 0.972 | 0.977 |
| $\sigma^2$ | 0.096 | 0.086 | 0.110 | 0.108 | 0.097 | 0.119 |
| $\phi$ | 0.465 | 0.421 | 0.515 | 0.468 | 0.440 | 0.491 |

| | Dividend Yield | | |
|---|---|---|---|
| $\theta$ | $\hat{\theta}$ | $q_{0.025}$ | $q_{0.975}$ |
| $\alpha_1$ | -5.617 | -5.953 | -5.234 |
| $\alpha_2$ | -5.117 | -5.472 | -4.793 |
| $p_{12}$ | 0.011 | 0.005 | 0.019 |
| $p_{22}$ | 0.989 | 0.983 | 0.994 |
| $\sigma^2$ | 0.094 | 0.059 | 0.136 |
| $\phi$ | 0.457 | 0.417 | 0.497 |

| | Industrial Production | | | Consumption Expenditure | | |
|---|---|---|---|---|---|---|
| $\theta$ | $\hat{\theta}$ | $q_{0.025}$ | $q_{0.975}$ | $\hat{\theta}$ | $q_{0.025}$ | $q_{0.975}$ |
| $\alpha_1$ | -4.911 | -5.025 | -4.788 | -0.898 | -0.928 | -0.871 |
| $\alpha_2$ | -4.228 | -4.400 | -4.073 | -0.307 | -0.349 | -0.260 |
| $p_{12}$ | 0.020 | 0.017 | 0.023 | 0.009 | 0.008 | 0.010 |
| $p_{22}$ | 0.999 | 0.988 | 0.991 | 0.979 | 0.975 | 0.983 |
| $\sigma^2$ | 0.369 | 0.318 | 0.421 | 0.204 | 0.169 | 0.244 |
| $\phi$ | 0.498 | 0.471 | 0.525 | 0.531 | 0.427 | 0.627 |

| | Non-Resid. Investment | | | Output Gap | | |
|---|---|---|---|---|---|---|
| $\theta$ | $\hat{\theta}$ | $q_{0.025}$ | $q_{0.975}$ | $\hat{\theta}$ | $q_{0.025}$ | $q_{0.975}$ |
| $\alpha_1$ | 0.626 | 0.535 | 0.736 | -0.803 | -0.999 | -0.606 |
| $\alpha_2$ | 0.783 | 0.624 | 0.959 | 0.107 | -0.080 | 0.304 |
| $p_{12}$ | 0.019 | 0.154 | 0.023 | 0.020 | 0.016 | 0.024 |
| $p_{22}$ | 0.980 | 0.976 | 0.983 | 0.989 | 0.987 | 0.991 |
| $\sigma^2$ | 0.124 | 0.101 | 0.143 | 0.201 | 0.166 | 0.241 |
| $\phi$ | 0.579 | 0.469 | 0.687 | 0.495 | 0.310 | 0.677 |

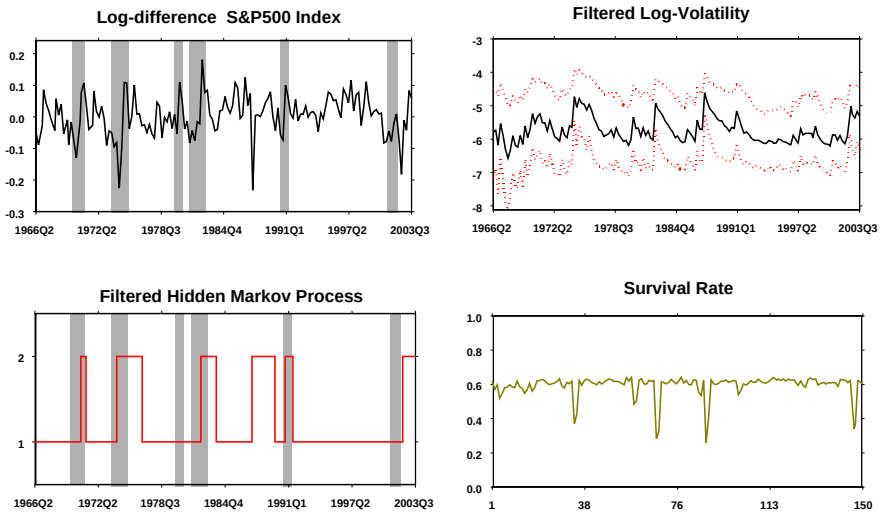


Figure 5: Log-differences of the S&P500 price index; sequentially filtered log-volatility with 0.025 and 0.975 quantiles (dotted lines); filtered volatility regimes and survival rate of the particle set, over the sample of $T = 150$ observations.

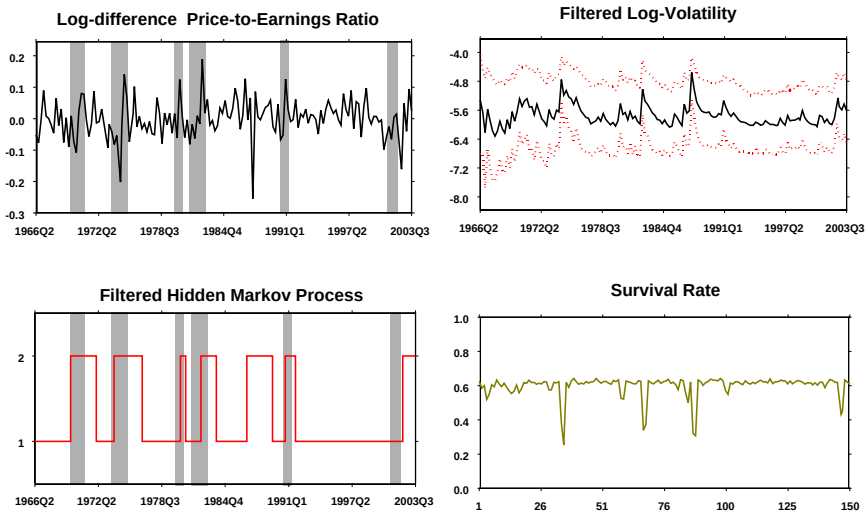


Figure 6: Log-differences of the Price-to-Earnings series; sequentially filtered log-volatility with 0.025 and 0.975 quantiles (dotted lines); filtered volatility regime and survival rate of the particle set, over the sample of $T = 150$ observations.

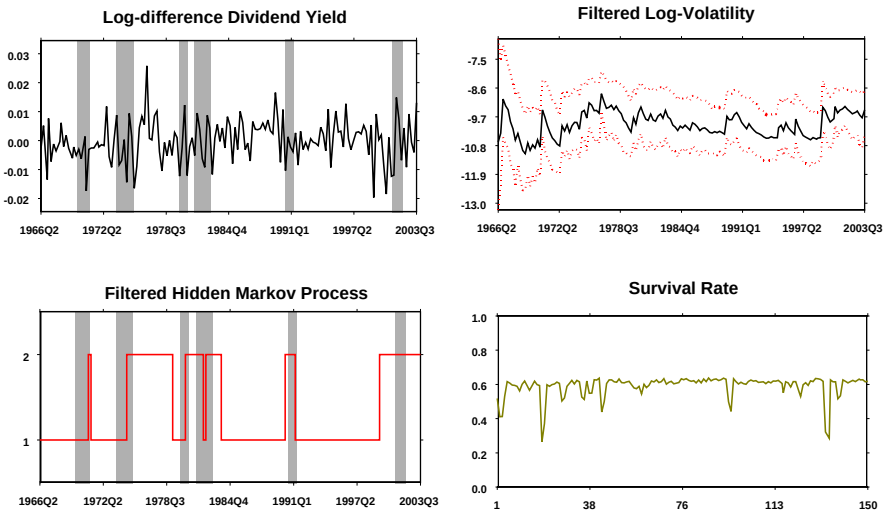


Figure 7: Log-differences of the Dividend Yield series; sequentially filtered log-volatility with 0.025 and 0.975 quantiles (dotted lines); filtered volatility regime and survival rate of the particle set, over the sample of $T = 150$ observations.

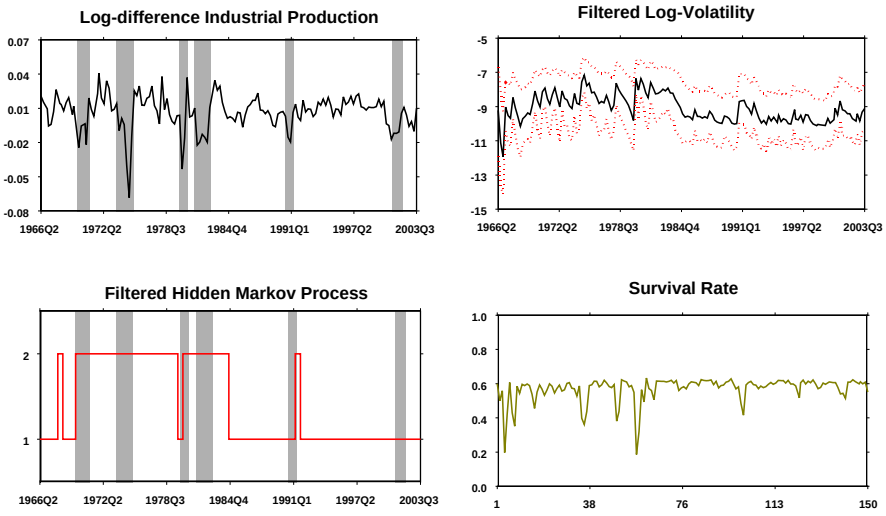


Figure 8: Log-differences of the Industrial Production index; sequentially filtered log-volatility with 0.025 and 0.975 quantiles (dotted lines); filtered volatility regime and survival rate of the particle set, over the sample of $T = 150$ observations.

part of the sample. Indeed, the estimated MSSV model shows that YGAP behaved according to a persistent high volatility regime essentially until 1984. Afterwards, it switched to a persistent low-volatility regime, with brief spells of high volatility around the 1990-1991 and 2001 contractions. Thus, our evidence says that YGAP volatility has followed a pattern that is somewhat intermediate between those of the other business cycle indicators, but overall closer to that of IP and NRI.

We can now draw some conclusions about the main objective of this exercise, namely, the investigation of similarities in the filtered volatilities of macroeconomic and stock market aggregate variables. Our indicators of economic activity follow a low-volatility regime for most of the second part of the sample. In particular, persistently low volatility characterizes industrial production, non-residential investment and the output gap; personal consumption expenditure switches to low volatility too, but with a lag relative to the other series. Our MSSV estimates detect essentially the same for our stock market indicators: the market index and its price-earnings and dividend-price ratios all switch to low-volatility around 1991. This switch appears to be persistent, as reversions to high volatility occurred only after the longest period of low volatility in our sample.

The estimates in Table 1 complement the graphical evidence we have just commented. In detail, the estimated values of $\alpha_1$ and $\alpha_2$, which provide the mean level of log-volatility of the series under the regime of low and high volatility, respectively, confirm that PCE, NRI and YGAP, that is, three out of the four macroeconomic variables, have the highest mean value of volatility. As to the values of the transition probabilities, we remind that they represent an indirect measure of persistence of each volatility regime. We observe that $p_{12}$, the probability of the variable switching from a low- to a high-volatility regime, is marginally higher for IP and YGAP, and smaller for S&P and DP. On the other hand, $p_{22}$, the probability of the variable staying in a high-volatility regime, is highest for S&P and YGAP, and lowest for PE and PCE[14]. Finally, $(1 - \phi)$ accounts for the speed of reversion of log-volatility to its mean value. We obtain the highest values for the $\phi$s of NRI and PCE, which therefore tend to revert to their mean volatility levels much more slowly than all other series. Stock market variables are, as expected, the quickest to revert to their mean values of log-volatility. In our approach all estimated parameters are time-varying. This allows us to gauge the stability of our estimated models: graphs in Appendix D confirm that estimated parameters are overall stable.

Figure 12 makes easier to summarize the volatility profiles of two pairs of variables. The top panel shows the volatility regimes for S&P and YGAP, the lower panel that for PE and IP. Both charts highlight that all series have

[14]We remind that $p_{11} = (1 - p_{12})$ is the probability of log-volatility remaining in the low-volatility regime, whereas $p_{21} = (1 - p_{22})$ is the probability of log-volatility migrating from a high to a low-volatility regime.

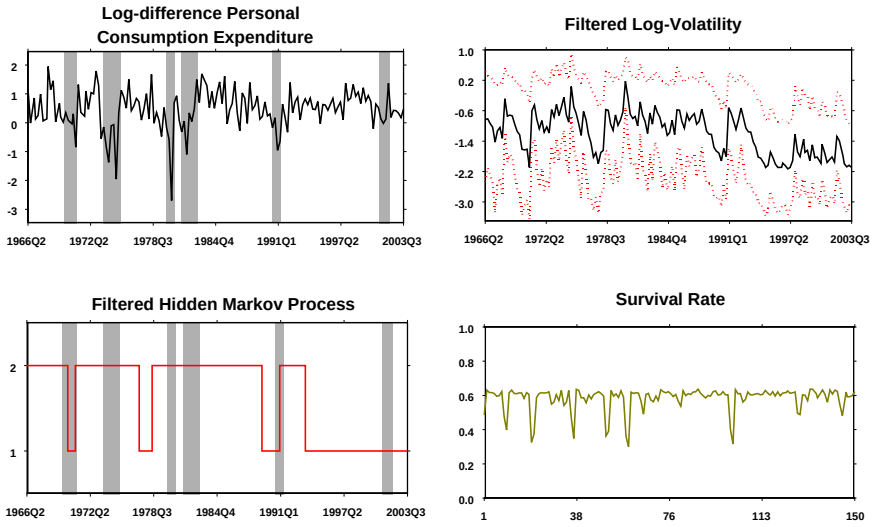


Figure 9: Log-differences of the Personal Consumption Expenditure per capita; sequentially filtered log-volatility with 0.025 and 0.975 quantiles (dotted lines); filtered volatility regime and survival rate of the particle set, over the sample of $T = 150$ observations.

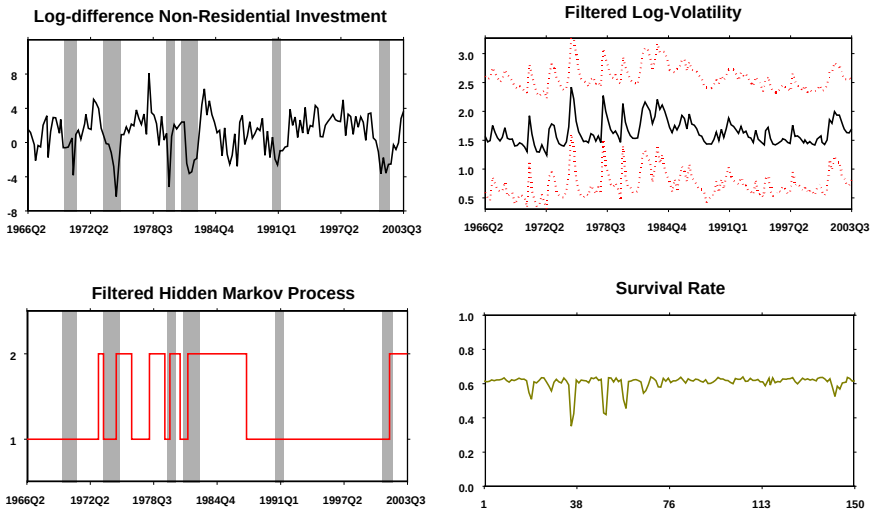


Figure 10: Log-differences of the non-residential Investments; sequentially filtered log-volatility with 0.025 and 0.975 quantiles (dotted lines); filtered volatility regime and survival rate of the particle set, over the sample of $T = 150$ observations.

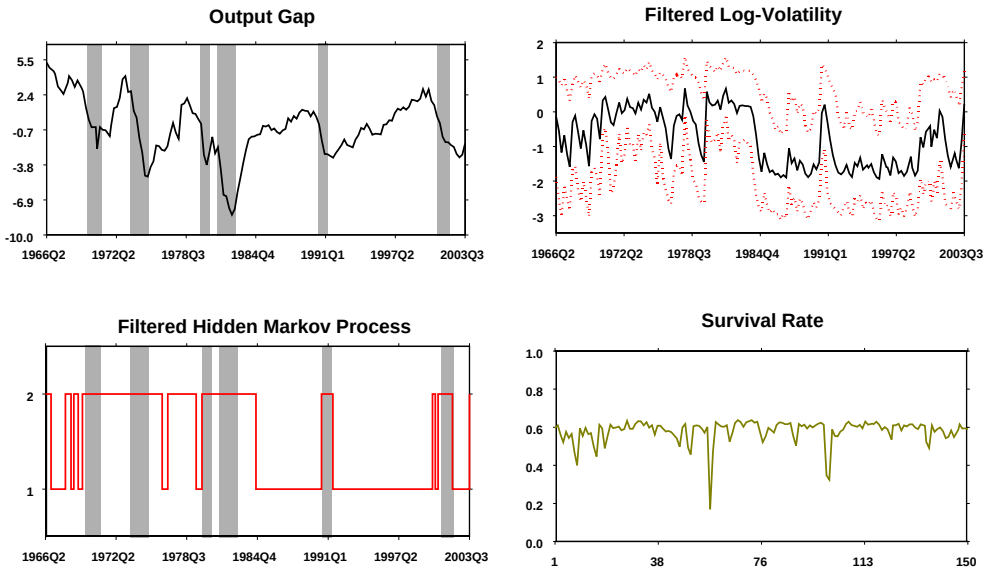


Figure 11: Output Gap; sequentially filtered log-volatility with 0.025 and 0.975 quantiles (dotted lines); filtered volatility regime and survival rate of the particle set, over the sample of $T = 150$ observations.

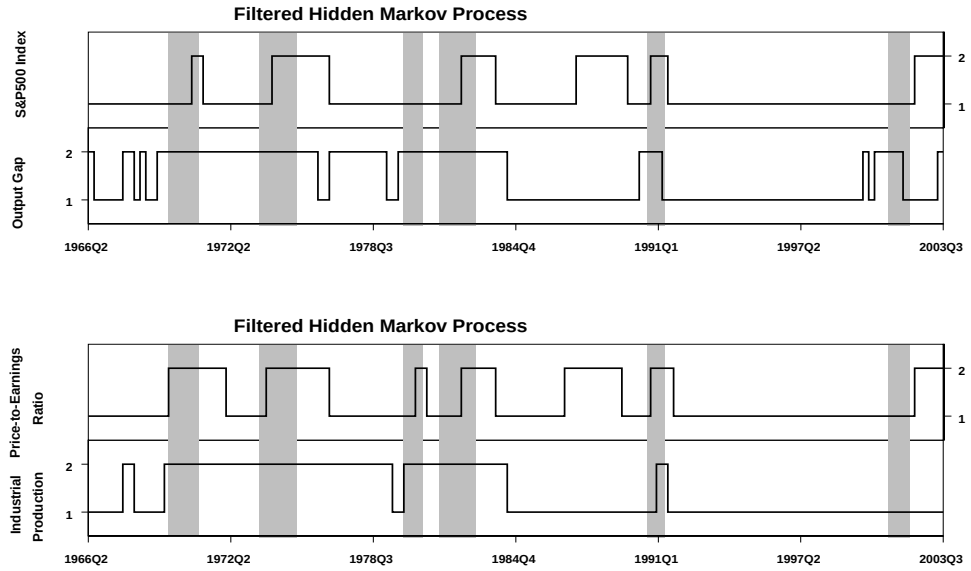


Figure 12: Filtered volatility regimes for selected variables.

switched to low volatility over the last half of the sample. In particular, the shift seems to have occurred in the business cycle variables first, and then in the stock market valuations. A natural development of this work, one which is needed before we can draw firmer conclusions on the issue at hand, is an extension to a multivariate setting. Such framework would enable to deal with causality and make more robust inference on the timing of the switches between volatility regimes.

# 5 Conclusions

In this work univariate Markov switching stochastic volatility models are estimated to extract the latent volatility regimes processes for some macroeconomic and aggregate stock market variables. We make Bayesian inference and propose a particle filter framework in order to jointly estimate the parameters and the hidden states of the dynamic model. The view we put to test is that the widely observed decline in US business cycle volatility translated into a similar long-term decrease in stock market volatility. According to this view, the relentless rise of stock prices in the late '90s might be due to a reduction of the equity premium, in turn generated by a decrease of broad macroeconomic risk, as represented by the noticed decline in business cycle volatility.

The evidence uncovered by our univariate estimates supports a parallel decline of macroeconomic and stock market variability, with the switch to low volatility occurring first in the business cycle indicators. Of course, *post hoc ergo propter hoc* (or *cum hoc ergo propter hoc*, for that matter) is a common fallacy and we certainly do not want to incur in it, but the findings of our empirical exercise are not inconsistent with the above view.

## Appendix A - Data

Stock market data are from Robert Shiller's website. His data set consists of monthly stock price, dividends, and earnings data, and the consumer price index (to allow conversion to real values), all starting January 1871. Monthly dividend and earnings data are computed from the S&P four-quarter tools for the quarter since 1926, with linear interpolation to monthly figures. Stock price data are monthly averages of daily closing prices.

Personal Consumption Expenditures and Real Gross Domestic Product data are from the U.S. Department of Commerce, Bureau of Economic Analysis, and are quarterly, seasonally adjusted observations. Real Potential Gross Domestic Product is from the U.S. Congress, Congressional Budget Office, and is a quarterly, seasonally adjusted series, in billions of chained 2000 Dollars. The output gap is calculated as the (log) difference between Real Gross Domestic Product and Real Potential Gross Domestic Product. Personal Consumption Expenditure is converted to real values using the Personal Consumption Expenditures Chain-type Price Index, quarterly, seasonally adjusted, index 2000=100, taken from the U.S. Department of Commerce, Bureau of Economic Analysis.

Real Private Non-Residential Fixed Investment and Real Residential Fixed Investment are from NIPA Tables, Bureau of Economic Analysis, and are quarterly, seasonally adjusted observations, in billions of chained 2000 dollars.

We converted all monthly data into quarterly averages.

In many charts vertical shaded areas represent the contraction phases of the US GDP as detected by Business Cycle Dating Committee of the NBER. The contraction starts at the peak of the cycle and ends at the through. In particular, our sample covers the following GDP contraction periods (the corresponding quarter is in parenthesis) that are of interest in our study: December 1969 (IV)-November 1970 (IV), November 1973 (IV)-March 1975 (I), January 1980 (I)-July 1980 (III), July 1981 (III)-November 1982 (IV), July 1990 (III)-March 1991 (I) and March 2001 (I)-November 2001 (IV).

## Appendix B - A Regularised Filtering Algorithm

In the following we give the multiple-bandwidth kernel regularised particle filter for the Markov Switching Stochastic Volatility model.

**Algorithm 1.** ***Kernel-Regularised Particle Filter***

*Given an initial set of particles* $\{x_t^i, s_t^i, \theta_t^i, \lambda_t^i, w_t^i\}_{i=1}^N$*:*

1. *Compute* $V_t = \sum_{i=1}^N (\theta_t^i - \bar{\theta}_t)(\theta_t^i - \bar{\theta}_t)' w_t^i$ *and* $\bar{\theta}_t = \sum_{i=1}^N \theta_t^i w_t^i$
2. *For* $i = 1, \ldots, N$ *and with* $a$ *and* $b$ *tuning parameters, calculate:*
   (a) $\tilde{S}_{t+1}^i = \arg\max_{l \in \{1,2\}} \mathbb{P}(s_{t+1} = l | s_t = s_t^i)$
   (b) $\tilde{X}_{t+1}^i = \alpha_t^i(\tilde{S}_{t+1}^i) + \phi_t^i x_t^i$
   (c) $\tilde{\theta}_t^i = a(\lambda_t^i)\theta_t^i + (1 - a(\lambda_t^i))\bar{\theta}_t$, *where* $\tilde{\theta} = (\tilde{\alpha}_{1,2}, \tilde{\phi}, \tilde{\sigma}, \tilde{\rho}, \tilde{\mu}, \tilde{p}_{11}, \tilde{p}_{22})$
3. *For* $i = 1, \ldots, N$*:*
   (a) *Simulate* $k^i$ *from* $\sum_{k=1}^N \mathcal{N}(y_{t+1}; \tilde{\mu}_t^k + \tilde{\rho}_t^k y_t, e^{\tilde{X}_{t+1}^k}) w_t^k \delta_{\{k\}}(dk^i)$
   (b) *Simulate* $s_{t+1}^i$ *from* $\mathbb{P}(s_{t+1}^i = l | s_t^{k^i})$ *with* $l \in \{1, 2\}$
   (c) *Simulate* $x_{t+1}^i$ *from* $\mathcal{N}(x_{t+1}; \alpha_t^{k^i}(s_{t+1}^i) + \phi_t^{k^i} x_t^{k^i}, \sigma_t^{k^i})$
   (d) *Simulate* $\theta_{t+1}^i$ *from* $\mathcal{N}_{n_\theta}(\theta_{t+1}; \tilde{\theta}_t^{k^i}, b^2(\lambda_t^{k^i}) V_t)$
4. *Update:* $\tilde{w}_{t+1}^i \propto \mathcal{N}(y_{t+1}; \tilde{\mu}_t^{k^i} + \tilde{\rho}_t^{k^i} y_t, e^{x_{t+1}^i}) / \mathcal{N}(y_{t+1}; \tilde{\mu}_t^{k^i} + \tilde{\rho}_t^{k^i} y_t, e^{\tilde{X}_{t+1}^{k^i}})$
5. *Normalize:* $w_{t+1}^i = \tilde{w}_{t+1}^i \, (\sum_{i=1}^N \tilde{w}_{t+1}^i)^{-1}$.

## Appendix C - Convergence Issues

Let $K$ be a Gaussian kernel and $f$ a density function both defined on $\mathbb{R}^d$. Note that $\int K(y)dy = 1$. We denote by $K_h(y) = h^{-d}K(y/h)$ a Gaussian kernel with *bandwidth* $h$ and by $(f * K_h)(x) = \int f(x-y)K_h(y)dy$ the convolution product. Let us define the following simple multiple-bandwidth kernel estimator.

**Definition 5.1.** *(Multiple-bandwidth estimator)*
*Let $X_1, \ldots, X_N$ be a sequence of i.i.d. random vectors with density $f$. The Multiple-bandwidth estimator of the density $f$ is*

$$f_N(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h_i^d}K(h_i^{-1}(x - X_i)). \tag{37}$$

*where $\{h_i\}_{i=1}^N$ is a sequence of positive numbers (bandwidths).*

Note that the proposed sample-point estimator is positive and integrates to one. Thus it is a density function and the Glick's theorem applies to its. The theorem gives a relation between a.s.-convergence and $L_1$-convergence for functional estimators.

**Theorem 5.1.** *(Glick's theorem)*
*Let $\{f_N(x)\}$ be a sequence of probability density functions on $\mathbb{R}^d$, which are measurable functions of $x$ and of $X_1, \ldots, X_N$ and such that $f_N \xrightarrow[N\to\infty]{a.s.} f$ almost everywhere in $x$. Then*

$$\int_{\mathbb{R}^d} |f_N(x) - f(x)|dx \xrightarrow[N\to\infty]{L_1} 0. \tag{38}$$

*Proof.* See Glick (1974). □

The following lemma is a useful preliminary result before showing some convergence results for the proposed multi-scale density estimator.

**Lemma 5.1.** *Let $K$ be a Gaussian kernel on $\mathbb{R}^d$, $f \in L^1(\mathbb{R}^d)$ a density function and $\{h_i\}_{i=1}^N$ a sequence of positive numbers (scale factors), which takes values in the interval $[h_{N,\min}, h_{N,\max}]$. Let $h_{N,\max}$ satisfy*

$$\lim_{N\to\infty} h_{N,\max} = 0 \tag{39}$$

*then*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(f * K_{h_i})(x) = f(x), \quad \forall x \in \mathbb{R}^d \tag{40}$$

*and*

$$\lim_{N\to\infty}\sup_x \left|\frac{1}{N}\sum_{i=1}^{N}(f * K_{h_i})(x) - f(x)\right| = 0. \tag{41}$$

*Proof.* This result extends the Bochner's lemma to the case of a variable-bandwidth kernel. Thus to prove the theorem we follow the technique used in Bosq and Lecoutre (1987) (p. 61). Let $\delta > 0$, then

$$\left|\frac{1}{N}\sum_{i=1}^{N}(f * K_{h_i})(x) - f(x)\right| \leq \tag{42}$$

$$\leq \frac{1}{N}\sum_{i=1}^{N}\int_{\|x_i\|\leq\delta} |(f(x-x_i)-f(x))K_{h_i}(x_i)|\,dx_i +$$

$$+\frac{1}{N}\sum_{i=1}^{N}\int_{\|x_i\|>\delta} |(f(x-x_i)-f(x))K_{h_i}(x_i)|\,dx_i$$

$$\leq \frac{1}{N}\sum_{i=1}^{N}\sup_{\|x_i\|\leq\delta}|(f(x-x_i)-f(x))|\int_{\|x_i\|\leq\delta}|K_{h_i}(x_i)|\,dx_i +$$

$$+\frac{1}{N}\sum_{i=1}^{N}\int_{\|x_i\|>\delta}\frac{|f(x-x_i)|}{\|\,x_i\,\|^d}\,\|\,\frac{x_i}{h_i}\,\|^d\,|K(x_i/h_i)|\,dx_i +$$

$$+\frac{1}{N}\sum_{i=1}^{N}|f(x)|\int_{\|x_i\|>\delta/h_i}|K(z_i)|\,dz_i$$

$$\leq \frac{1}{N}\sum_{i=1}^{N}\sup_{\|x_i\|\leq\delta}|(f(x-x_i)-f(x))|\int_{\mathbb{R}^d}|K(z_i)|\,dz_i +$$

$$+\frac{1}{N}\sum_{i=1}^{N}\delta^{-d}\int_{\mathbb{R}^d}|f(x-z_ih_i)|\sup_{\|z_i\|>\delta/h_i}\|\,z_i\,\|^d\,|K(z_i)|\,dz_i +$$

$$+\frac{1}{N}\sum_{i=1}^{N}|f(x)|\int_{\|z_i\|>\delta/h_i}|K(z_i)|\,dz_i$$

Let $\delta$ be fixed and $N\to\infty$ then $h_{N,\max}\to 0 \Rightarrow h_i\to 0$, $\forall\, i$ and the last two summations tend to zero, because for Gaussian kernels $\|\,z\,\|^d\,|K(z)|\to 0$ as $\|\,z\,\|^d\to\infty$. Now let $\delta$ tend to zero, also the first summation equals to zero.

The second part of the lemma follows from the inequalities

$$\sup_x\left|\frac{1}{N}\sum_{i=1}^{N}(f*K_{h_i})(x)-f(x)\right| \leq \frac{1}{N}\sum_{i=1}^{N}\sup_x|(f*K_{h_i})(x)-f(x)| \tag{43}$$

$$\leq \frac{1}{N}\sum_{i=1}^{N}\left\{\sup_x\sup_{\|x_i\|\leq\delta}|(f(x-x_i)-f(x))|\int|K(z_i)|\,dz_i\right\}+$$

$$+\frac{1}{N}\sum_{i=1}^{N}\left\{2\sup_x|f(x)|\int_{\|z_i\|>\delta/h_i}|K(z_i)|\,dz_i\right\}$$

which tends to zero due to result in the first part of the lemma.

□

**Theorem 5.2.** *(Quadratic-mean convergence)*
*Let $K$ be a Gaussian kernel on $\mathbb{R}^d$, $f \in L^1(\mathbb{R}^d)$ a density, $\{x_i\}_{i=1}^N$ i.i.d. samples with common density $f$ and $\{h_i\}_{i=1}^N$ a sequence of positive numbers (bandwidths) with values in the interval $[h_{N,\min}, h_{N,\max}]$. Let the bounds satisfy*

$$\lim_{N\to\infty} h_{N,\max} = 0\,, \quad \lim_{N\to\infty} (h_{N,\min})^d N = \infty \tag{44}$$

*then the sample-point estimators, $f_N$, converges in $L_2$ to the true density*

$$f_N(x) \xrightarrow[N\to\infty]{L_2} f(x), \quad \text{a.e. in } x. \tag{45}$$

*Proof.* Take the expectation of the functional estimator with respect to the sequence of i.i.d. random vectors $\{X_i\}$

$$\begin{aligned}
&\mathbb{E}\left(f_N(x)\right) = \qquad (46)\\
=\;& \int_{\mathbb{R}^{Nd}} \left(\frac{1}{N}\sum_{i=1}^N \frac{1}{h_i^d} K(h_i^{-1}(x-x_i))\right) f(x_1)\dots f(x_N)dx_1\dots dx_N\\
=\;& \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} \frac{1}{h_i^d} K(h_i^{-1}(x-x_i)) f(x_i)dx_i\\
=\;& \frac{1}{N}\sum_{i=1}^N (f * K_{h_i})(x)
\end{aligned}$$

which is the quantity studied in Lemma 5.1. Now the quadratic error decomposes as follows

$$\begin{aligned}
&\mathbb{E}(f_N(x) - f(x))^2 = \qquad (47)\\
=\;& \mathbb{E}[f_N(x) - \mathbb{E}(f_N(x))]^2 + [\mathbb{E}(f_N(x)) - f(x)]^2\\
=\;& \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^N \frac{1}{h_i^d} K(h_i^{-1}(x-x_i)) - \frac{1}{N}\sum_{i=1}^N (f * K_{h_i})(x)\right)^2 +\\
&+[\mathbb{E}(f_N(x)) - f(x)]^2\\
=\;& \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^N \frac{1}{h_i^d} K(h_i^{-1}(x-x_i))\right)^2 - \left(\frac{1}{N}\sum_{i=1}^N (f * K_{h_i})(x)\right)^2 +\\
&+[\mathbb{E}(f_N(x)) - f(x)]^2
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N^2}\sum_{i=1}^{N}\int \frac{1}{h_i^{2d}}K^2(h_i^{-1}(x-x_i))f(x_i)dx_i \\
&\quad + \frac{1}{N^2}\sum_{i\neq j}\int \frac{1}{h_i^d h_j^d}K(h_i^{-1}(x-x_i))K(h_j^{-1}(x-x_j))f(x_j)f(x_i)dx_j dx_i \\
&\quad - \left(\frac{1}{N}\sum_{i=1}^{N}(f*K_{h_i})(x)\right)^2 + [\mathbb{E}(f_N(x)) - f(x)]^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{h_i^d}(f*K_{h_i}^2)(x) - \frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{h_i^{2d}}\left(\int K(h_i^{-1}(x-x_i))f(x_i)\right)^2 dx_i + \\
&\quad + [\mathbb{E}(f_N(x)) - f(x)]^2 \\
&\leq \frac{\int K^2}{N h_{N,\min}^d}\left(\frac{1}{N}\sum_{i=1}^{N}(f*\tilde{K}_{h_i}^2)(x)\right) - \frac{1}{N^2}\left(\sum_{i=1}^{N}\mathbb{E}^2\left(K_{h_i}(x-x_i)\right)\right) + \\
&\quad + [\mathbb{E}(f_N(x)) - f(x)]^2
\end{aligned}
$$

where $\tilde{K}_{h_i}^2(x) = K_{h_i}^2(x)/\int K^2$ is a normalized kernel, integrating to one.

Lemma 5.1 and the assumption: $h_{N,\max} \to 0$ as $N \to \infty$, insure that the first term in parenthesis converges to $f(x)$ and the last term, converges to zero. Each element of the sum in the middle is bounded. The two elements tend to zero providing that $\lim_{N\to\infty}(h_{N,\min})^d N = \infty$.

□

The following theorem states the almost sure (a.s.) convergence of the proposed sample-point estimator.

**Theorem 5.3.** *(a.s. convergence)*
*Let $K$ be a Gaussian kernel on $\mathbb{R}^d$, $f \in L^1(\mathbb{R}^d)$ a density, $\{x_i\}_{i=1}^N$ i.i.d. samples from $f$ and $\{h_i\}_{i=1}^N$ a sequence of positive numbers (bandwidths) with values in the interval $[h_{N,\min}, h_{N,\max}]$. Let the bounds satisfy*

$$
\lim_{N\to\infty} h_{N,\max} = 0\,, \quad \lim_{N\to\infty}\frac{(h_{N,\min})^d N}{\log N} = \infty \tag{48}
$$

*then the sample-point estimators, $f_N$, converges a.s. to the true density*

$$
f_N(x) \xrightarrow[N\to\infty]{a.s.} f(x) \tag{49}
$$

*Proof.* We follow Devroye and Wagner (1979) and apply the triangular inequality to decompose the estimation error into the variance and bias components.

$$
|f_N(x) - f(x)| \leq |f_N(x) - \mathbb{E}(f_N(x))| + |\mathbb{E}(f_N(x)) - f(x)|. \tag{50}
$$

We first consider the second term on the RHS of the inequality and show that by a straightforward application of the Lemma 5.1

$$\mathbb{E}(f_N(x)) \xrightarrow[N\to\infty]{a.s.} f(x).$$

The second result in Lemma 5.1 insures the pointwise convergence of $\mathbb{E}(f_N(x))$ to $f(x)$ uniformly in $x$ and implies

$$|\mathbb{E}(f_N(x)) - f(x)| = \left| \frac{1}{N} \sum_{i=1}^{N} (f * K_{h_i})(x) - f(x) \right| \xrightarrow[N\to\infty]{a.s.} 0 \tag{51}$$

providing that $h_{N,\max} \to 0$ as $N \to \infty$ and $\| z \|^d |K(z)| \to 0$ as $\| z \|^d \to \infty$.

Consider now the first term on the RHS of the inequality,

$$|f_N(x) - \mathbb{E}(f_N(x))| = \left| \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{h_i^d} K(h(x - x_i)) - (f * K_{h_i})(x) \right) \right|. \tag{52}$$

Let $M = \sup_z K(z)$ and note that

$$K_{h_i} \le h_i^{-d} M \le h_{N,\min}^{-d} M$$

and

$$\mathbb{E}(K_{h_i}^2) = h_i^{-d} (f * K_{h_i}^2)(x) \le h_{N,\min}^{-d} h_i^{-d} M (f * K_{h_i})(x).$$

By Bernstein-Fréchet's inequality[15] ,

$$\begin{aligned}
& P\left(|f_N(x) - \mathbb{E}(f_N(x))| \ge \varepsilon\right) \le \\
\le \;& 2\exp\left( -\frac{\varepsilon^2 N^2}{4h_{N,\min}^{-d} M \sum_i \mathbb{V}(K_{h_i}) + 2\varepsilon N h_{N,\min}^{-d} M} \right) \\
< \;& 2\exp\left( -\frac{\varepsilon^2 N^2 h_{N,\min}^d}{4M \sum_i \mathbb{E}(K_{h_i}) + 2\varepsilon N M} \right) \\
\le \;& 2\exp\left( -\frac{\varepsilon^2 N^2 h_{N,\min}^d}{4M \sum_i h_i^{-d} (f * K_{h_i})(x) + 2\varepsilon N M} \right) \\
= \;& 2\exp\left( -\frac{\varepsilon^2 N h_{N,\min}^d}{4M\mathbb{E}(f_N(x)) + 2\varepsilon M} \right) \le \exp\left(\alpha N h_{N,\min}^d\right)
\end{aligned} \tag{53}$$

which is bounded a.e. in $x$ and for all $N$, due to the a.s. convergence of $\mathbb{E}(f_N(x))$ to $f(x)$. The above inequality insures the almost complete (a.c.)

[15] See Bosq and Lecoutre (1987) (Th. I.2, p. 41), with $p = 3$, and $\mathbb{E}|K_{h_i} - \mathbb{E}(K_{h_i})|^3 \le h_{N,\min}^{-d} M \mathbb{V}(K_{h_i})$.

convergence and consequently the a.s. convergence (see Bosq and Lecoutre (1987)) of the estimator

$$f_N(x) \xrightarrow[N\to\infty]{a.c.} \mathbb{E}(f_N(x)),$$

providing that $\sum_{N=1}^{\infty} e^{-\alpha N h_{N,\min}^d} < \infty$. This condition is clearly implied by the assumption: $\lim_{N\to\infty} A_N = \infty$ where $A_N = N h_{N,\min}^d \log(N)^{-1}$. For an arbitrarily chosen $\eta > 0$ there exists $M$ s.t. for all $N > M$, $A_N > \eta$ and

$$\begin{aligned}
&\sum_{N=1}^{\infty} e^{-\alpha N h_{N,\min}^d} \sum_{N=1}^{\infty} e^{-\alpha \log(N) A_N} < \\
&< \sum_{N=1}^{M} e^{-\alpha \log(N) A_N} + \sum_{N=M+1}^{\infty} e^{-\alpha \log(N) \eta} \\
&< \sum_{N=1}^{M} e^{-\alpha \log(N) A_N} + \sum_{N=1}^{\infty} N^{-\alpha\eta} < \infty
\end{aligned} \tag{54}$$

choosing $\eta$ such that the series is convergent. □

**Theorem 5.4.** *($L_1$-convergence)*
*Let $K$ be a Gaussian kernel on $\mathbb{R}^d$, $f \in L^1(\mathbb{R}^d)$ a density, $\{x_i\}_{i=1}^N$ i.i.d. samples from $f$ and $\{h_i\}_{i=1}^N$ a sequence of positive numbers (bandwidths) with values in the interval $[h_{N,\min}, h_{N,\max}]$. Let the bounds satisfy*

$$\lim_{N\to\infty} h_{N,\max} = 0\,, \quad \lim_{N\to\infty} \frac{(h_{N,\min})^d N}{\log N} = \infty \tag{55}$$

*then the sample-point estimators, $f_N$, converges in $L_1$ to the true density*

$$f_N(x) \xrightarrow[N\to\infty]{L_1} f(x). \tag{56}$$

*Proof.* Note first that the assumptions of the theorem insure the a.s.-convergence (see Theorem (5.3)) of the multiple-bandwidth estimator. The a.s. convergence implies the $L_1$-convergence by the Glick's theorem. □

# Appendix D - Recursive Parameter Estimates

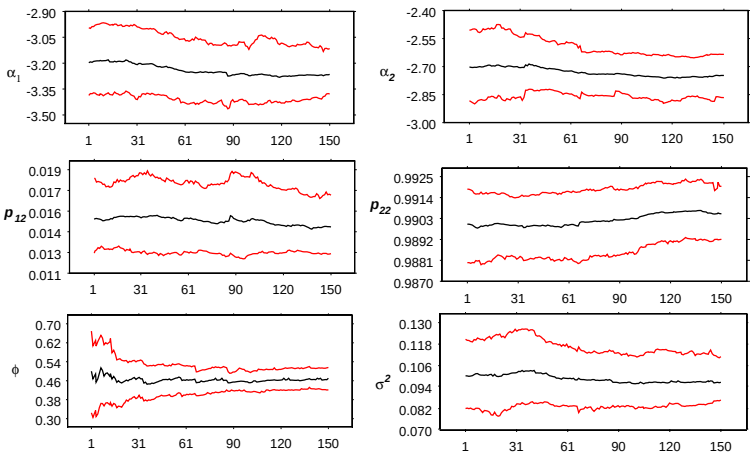


Figure 13: On-line parameter estimates for the log-returns on the S&P price index. Graphs show at each date the empirical mean and the quantiles at 0.025 and 0.975 for each parameter.

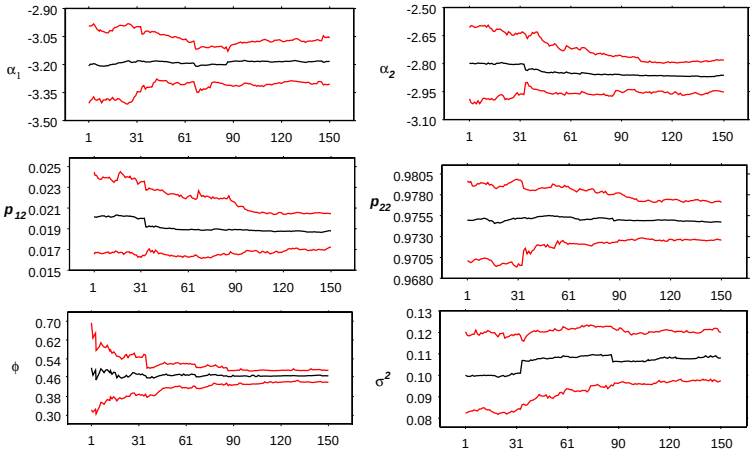


Figure 14: On-line parameter estimates for the Price-to-Earnings. Graphs show at each date the empirical mean and the quantiles at 0.025 and 0.975 for each parameter.

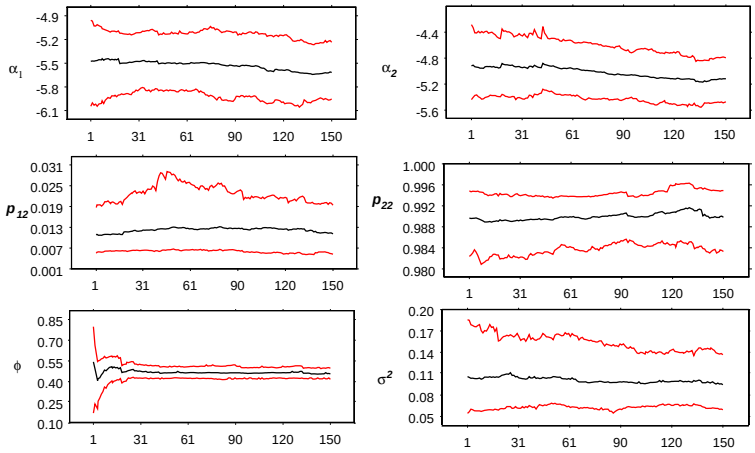


Figure 15: On-line parameter estimates for the Dividend Yields. Graphs show at each date the empirical mean and the quantiles at 0.025 and 0.975 for each parameter.

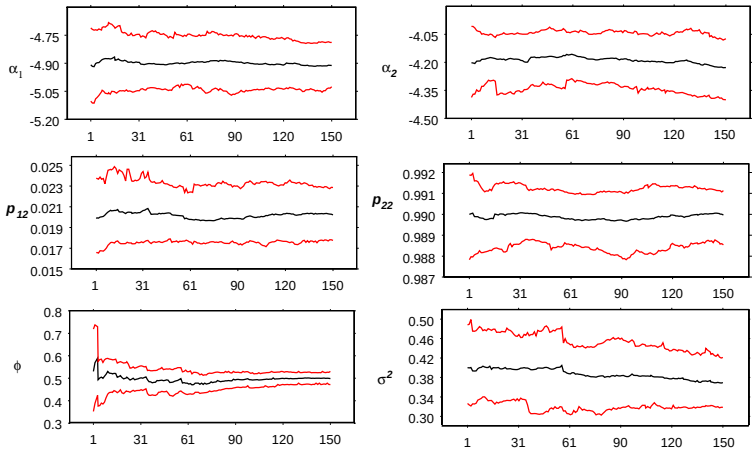


Figure 16: On-line parameter estimates for the log-returns on the Industrial Production index. Graphs show at each date the empirical mean and the quantiles at 0.025 and 0.975 for each parameter.

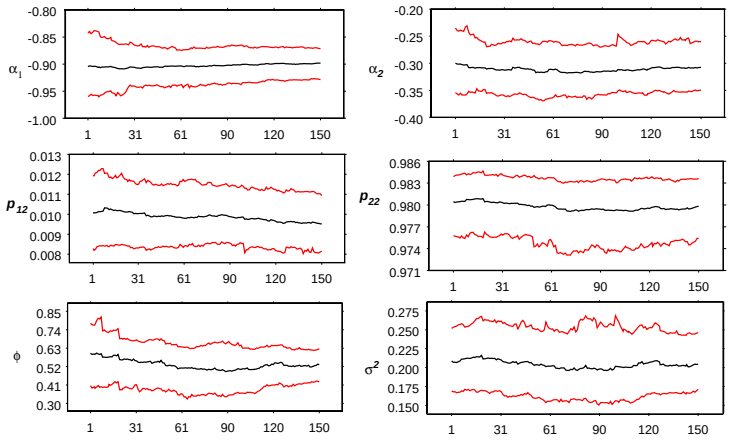


Figure 17: On-line parameter estimates for the Personal Consumption Expenditure. Graphs show at each date the empirical mean and the quantiles at 0.025 and 0.975 for each parameter.

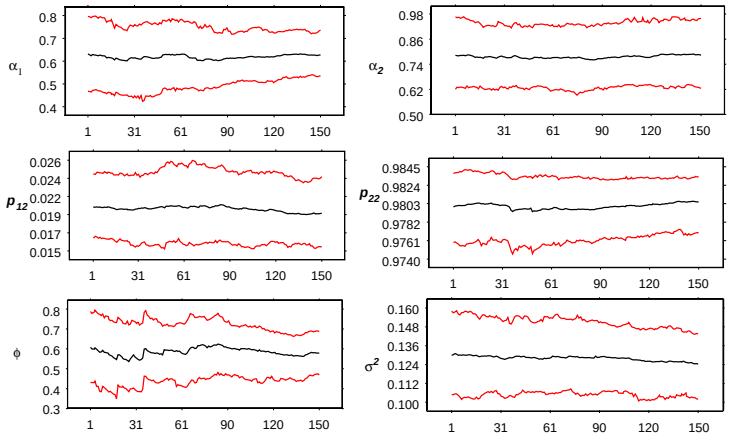


Figure 18: On-line parameter estimates for the non-residential investments series. Graphs show at each date the empirical mean and the quantiles at 0.025 and 0.975 for each parameter.

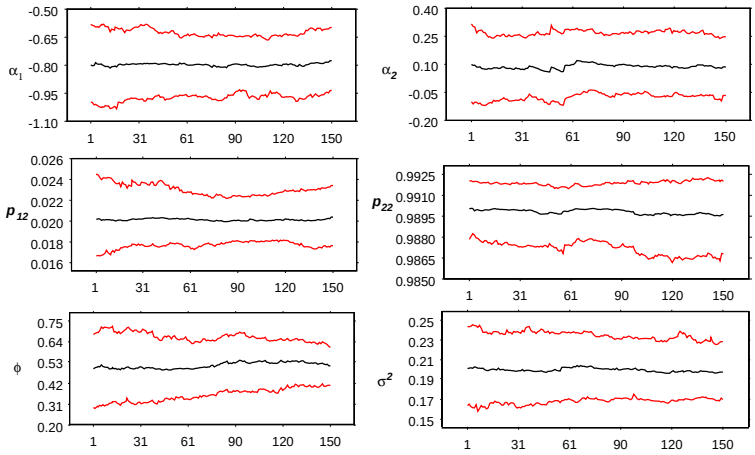


Figure 19: On-line parameter estimates for the output gap series. Graphs show at each date the empirical mean and the quantiles at 0.025 and 0.975 for each parameter.

## References


Arulampalam, S., S. Maskell, N. Gordon, and T. Clapp (2001): "A Tutorial on Particle Filters for On-line Nonlinear/Non-Gaussian Bayesian Tracking," Technical Report, QinetiQ Ltd., DSTO, Cambridge.

Bartoli, N., and P. Del Moral (2001): *Simulation et algorithmes stochastiques*. Cépaduès Éditions, Paris.

Bauwens, L., M. Lubrano, and J. F. Richard (1999): *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, New York.

Berzuini, C., N. G. Best, W. R. Gilks, and C. Larizza (1997): "Dynamic conditional independence models and Markov chain Monte Carlo methods," *Journal of the American Statistical Association*, 92(440), 1403–1441.

Berzuini, C., and W. R. Gilks (2001): "Following a moving average-Monte Carlo inference for dynamic Bayesian models," *Journal of Royal Statistical Society, B*, 63(1), 127–146.

Billio, M., R. Casarin, and D. Sartore (2004): "Bayesian Inference on Dynamic Models with Latent Factors," Official Monography, Edited by EUROSTAT.

Borio, C., and P. Lowe (2004): "Securing sustainable price stability: should credit come back from the wilderness?," BIS Working Papers N. 157.

Bosq, D., and J. P. Lecoutre (1987): *Théorie de l'estimation fonctionelle*. Economica, Paris.

Casarin, R. (2004): "Bayesian Inference for Markov Switching Stochastic Volatility Models," Cahier du CEREMADE N. 0414, Université Paris Dauphine.

Chopin, N. (2001): "Sequential inference and state number determination for discrete state-space models through particle filtering," Technical Report N. 34, CREST.

Chopin, N., and F. Pelgrin (2004): "Bayesian inference and state number determination for hidden Markov models: an application to the information content of the yield curve about inflation," *Journal of Econometrics*, 123(2), 327–344.

Clarida, R., J. Gali, and M. Gertler (1998): "Monetary policy rules in practice: some international evidence," *European Economic Review*, 42(6), 1033–1067.

Cochrane, J. H. (2005): *Asset Pricing (Revised edition)*. Princeton University Press, Princeton.

Crisan, D. (2001): "Particle filters, A theoretical perspective," in *Sequential Monte Carlo Methods in Practice*, ed. by A. Doucet, J. Freitas, and J. Gordon. Springer-Verlag, New York.

Crisan, D., and A. Doucet (2000): "Convergence of sequential Monte Carlo methods," Technical Report N. 381, CUED-F-INFENG.

——— (2002): "A survey of convergence results on particle filtering for practitioners," *IEEE Trans. Signal Processing*, 50, 736–746.

Del Moral, P. (2004): *Feynman-Kac Formulae*. Springer Verlag, New York.

Devroye, L. P., and T. J. Wagner (1979): "The $L_1$ Convergence of Kernel Density Estimates," *The Annals of Statistics*, 7(5), 1136–39.

Doucet, A., J. G. Freitas, and J. Gordon (2001): *Sequential Monte Carlo Methods in Practice*. Springer Verlag, New York.

Doucet, A., S. Godsill, and C. Andrieu (2000): "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, 10(3), 197–208.

Dynan, K. E., D. W. Elmendorf, and D. E. Sichel (2006): "Can financial innovation help to explain the reduced volatility of economic activity?," *Journal of Monetary Economics*, 53, 123–150.

Ferson, W. (2003): "Test of Multi-Factor Pricing Models, Volatility, and Portfolio Performance," in *Handbook of the Economics of Finance, 1B*, ed. by G. M. Constantinides, M. Harris, and R. M. Stulz. Elsevier, North-Holland.

Gilchrist, S., C. P. Himmelberg, and G. Huberman (2005): "Do Stock Price Bubbles Influence Corporate Investment?," *Journal of Monetary Economics*, 52, 805–827.

Glick, N. (1974): "Consistency conditions for probability estimators and integrals of density estimators," *Utilitas Mathematica*, 6, 61–74.

Gordon, N., D. Salmond, and A. F. M. Smith (1993): "Novel Approach to Nonlinear and Non-Gaussian Bayesian State Estimation," *IEE Proceedings-F*, 140, 107–113.

Gordon, R. J. (2005): "What caused the decline in U.S. business cycle volatility?," NBER Working Paper N. 11777.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384.

Harrison, J., and M. West (1997): *Bayesian Forecasting and Dynamic Models, 2nd Ed.* Springer Verlag, New York.

Harvey, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

Kalman, R. E., and R. S. Bucy (1960): "New results in linear filtering and prediction problems," *Transaction of the ASME-Journal of Basic Engineering, Series D*, 83, 95–108.

Kim, C. J., and C. R. Nelson (1999): *State-Space Models with Regime Switching*. MIT press, Cambridge.

Kitagawa, G. (1998): "A self-organizing state space model," *Journal of the American Statistical Association*, 93, 1203–1215.

Lettau, M., and S. C. Ludvigson (2001): "Consumption, aggregate wealth, and expected stock returns," *Journal of Finance*, 3, 815–849.

——— (2004): "Understanding Trend and Cycle in Asset Value: Reevaluating the Wealth Effect on Consumption," *American Economic Review*, 94, 276–299.

Lettau, M., S. C. Ludvigson, and J. A. Wachter (2005): "The declining equity premium: what role does macroeconomic risk play?," Working paper, Stern School of Business, NYU.

Liu, J., and M. West (2001): "Combined Parameter and State Estimation in Simulation Based Filtering," in *Sequential Monte Carlo Methods in Practice*, ed. by A. Doucet, J. Freitas, and J. Gordon. Springer-Verlag, New York.

Liu, J. S., and R. Chen (1998): "Sequential Monte Carlo Methods for Dynamical System," *Journal of the American Statistical Association*, 93(443), 1032–1044.

Lopes, H. F., and C. Marigno (2001): "A particle filter algorithm for the Markov switching stochastic volatility model," Working paper, University of Rio de Janeiro.

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

Menzly, L., T. Santos, and P. Veronesi (2004): "Understanding predictability," *Journal of Political Economy*, 112(1), 1–47.

Musso, C., N. Oudjane, and F. LeGland (2001): "Improving Regularised Particle Filters," in *Sequential Monte Carlo in Practice*, ed. by A. Doucet, J. Freitas, and J. Gordon. Springer Verlag, New York.

Perpiñán, M. A. C. (2001): "Continuous latent variable models for dimensionality reduction and sequential data reconstruction," PhD Thesis in Computer Science, University of Sheffield.

Pitt, M., and N. Shephard (1999): "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94(446), 590–599.

Polson, N. G., J. R. Stroud, and P. Müller (2002): "Practical Filtering with sequential parameter learning," Tech. report, Graduate School of Business, University of Chicago.

——— (2003): "Practical Filtering for Stochastic Volatility Models," Tech. report, Graduate School of Business, University of Chicago.

Robert, C. P., and G. Casella (2004): *Monte Carlo Statistical Methods, 2nd ed.* Springer Verlag, New York.

Rossi, V. (2004): "Filtrage non linéaire par noyaux de convolution. Application à un procédé de dépollution biologique," Thèse du Doctorat en Sciénce, Ecole Nationale Superiore Agronomique de Montpellier.

Shiller, R. J. (2000): *Irrational Exuberance*. Princeton University Press, Princeton.

So, M. K. P., K. Lam, and W. K. Li (1998): "A stochastic volatility model with Markow switching," *Journal of Business & Economic Statistics*, 16, 244–253.

Stavropoulos, P., and D. M. Titterington (2001): "Improved Particle filters and smoothing," in *Sequential Monte Carlo Methods in Practice*, ed. by A. Doucet, J. Freitas, and J. Gordon. Springer-Verlag, New York.

Stock, J. H., and M. W. Watson (2002): "Has the Business Cycle Changed and Why?," in *NBER Macroeconomics Annual*, ed. by M. Gertler, and K. Rogoff. MIT Press, Cambridge.

Storvik, G. (2002): "Particle filters for state space models with the presence of unknown static parameters," *IEEE Trans. on Signal Processing*, 50, 281–289.

Villaverde, J. F., and J. F. Ramirez (2004a): "Estimating Nonlinear Dynamic Equilibrium Economies: A Likelihood Approach," Atlanta FED Working Paper N. 03.

——— (2004b): "Estimating Nonlinear Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood," Atlanta FED Working Paper N. 01.

Watson, J. (1994): "Business cycle durations and postwar stabilization of the U.S. economy," *American Economic Review*, 84, 24–46.

West, M. (1992): "Mixture models, Monte Carlo, Bayesian updating and dynamic models," *Computing Science and Statistics*, 24, 325–333.

——— (1993): "Approximating posterior distribution by mixtures," *Journal of Royal Statistical Society, B*, 55, 409–442.