Filtering via Taylor Series Approximations*

Alberto Romero

Sauder School of Business

University of British Columbia

This Version: September 2012

Abstract

I propose a nonlinear filtering method to estimate latent VAR processes based on high order Taylor approximations. I study three different applications from Finance and Economics and demonstrate that this filtering method is a good approach for state and parameter estimation. My estimation results show large differences between the state variable estimates obtained via high order Taylor Series filters and conventional methods such as the Extended Kalman Filter or the Unscented Kalman Filter. My results suggest that this filter may prove to be a good approach for a number of problems that involve nonlinear dynamic modeling in Finance and Economics.

JEL classification: C4; C32.

Keywords: Kalman Filter; Taylor series.

*I am grateful to Adlai Fisher and Lorenzo Garlappi for their guidance and supervision. I thank Nando de Freitas, Howard Kung, Georgios Skoulakis, Vincent Gregoire and Gonzalo Morales for their comments and suggestions. All errors are mine. Contact information: 2053 Main Mall, Vancouver BC, Canada V6T 1Z2; Email: `alberto.romero@sauder.ubc.ca`.

1 Introduction

Filtering is a statistical tool that recovers unobservable state variables using measurements that are observed with "noise". This technique has been the subject of considerable research interest during past decades due to the wide variety of applications in science and engineering which include satellite navigation systems, tumor identification, weather forecasting. Kalman (1960) proposed a well known solution to the linear filtering problem, the Kalman filter, that computes the estimates of the state of a system, given the set of observations available. It has been applied to problems in economics and finance where agents make decisions based on "noisy" information. Generalizations of the Kalman filter, commonly referred to as nonlinear filters, allow state variables to have a nonlinear relation with measurements or previous states; however, a complication occurs in this case, since the most common solutions for the linear filtering problem are not valid.

In this paper, I propose a nonlinear filtering method to estimate latent vector autorrecgressive (VAR) processes based on high order Taylor approximations. The method can be applied for both state and parameter inference, using standard Quasi-maximum likelihood techniques as in White (1982) and Bollerslev and Wooldridge (1992). Finally, to test the accuracy of the filter, I implement the filtering method with three different nonlinear models that have been studied in the Finance and Economics literature.

The first application is the Stochastic Volatility model by Andersen and Sørensen (1996), Andersen, Bollerslev, Diebold, and Ebens (2001), Andersen, Bollerslev, Diebold, and Labys (2003) and Broto and Ruiz (2004). I first study state and parameter estimation for the standard stochastic volatility model with simulated data. My findings suggest that the filtering method with Taylor series approximations is an alternative approach for both state and parameter inference. Finally, I estimate the parameters of an endowment process with stochastic volatility using vintage series of monthly consumption growth for US data and find evidence of stochastic volatility. Moreover, my results are comparable with the recent findings by Bidder and Smith (2011) and Ludvigson (2012) in that suggest that the stochastic volatility model is a good representation for consumption growth, as in the long-run risks literature (Bansal and Yaron (2004)).

In the second application, I analyze a nonlinear latent vector autoregressive (VAR) process studied in Brandt and Kang (2004) (BK hereafter) and recently used by Boguth, Carlson, Fisher, and Simutin (2011) in the conditional asset pricing literature. BK model the conditional mean and volatility of stock returns as a two dimensional latent VAR process. This approach has several advantages it guarantees positive risk premia and volatilities; eliminates the reliance of arbitrary conditioning variables for the construction of conditional moments and allows to study the contemporaneous and intertemporal relationships between expected returns and risk. I find fourth, the VAR representation captures different models.[1].

The last application is a stochastic general equilibrium model, which is particularly interesting as it shows that perturbation techniques that have been previously used to solve general equilibrium models[2] can be directly combined with nonlinear filtering for state and parameter estimation. Moreover, the filter with Taylor approximations may be another feasible approach for parameter inference of this type of models since a quasi-likelihood function can be constructed, avoiding Monte Carlo simulation procedures.

As for the nonlinear filtering literature, two widely used algorithms have been successfully applied: the Extended Kalman filter (Jazwinski (1970)) and the Unscented Kalman filter (Julier and Uhlmann (1997)). These approaches rely on first and second order approximations of the functions that characterize the nonlinear data generating process; however, if nonlinearities are significant enough, these filters do not provide efficient estimates and a number of biases arise.[3]

This paper extends these approaches by allowing an arbitrary order of approximation. The filter is based on the multivariate Taylor series expansions recently used in Savits (2006) and Garlappi and Skoulakis (2010) in two ways: first, for the computation of higher order derivatives of functions and second for the computation of higher order moments of normally distributed random variables. The nonlinear filtering technique that I propose overcomes most of the difficulties documented before in the filtering literature for a number of reasons: first, it allows for arbitrary nonlinearities in the data generating process; second, the accuracy level is exogenous and chosen by the researcher; third, the filtering calculations are potentially as efficient as the standard Kalman Filter since only function evaluations are required to calculate the recursions; and fourth, the precision of the filtered states increases as the degree of approximation increases. From an econometric point of view, the filters based on Taylor series approximations can be applied for inference purposes, since a quasi-likelihood function can be constructed from the filtering recursions, White (1982), and robust standard errors can be computed according to the method of Bollerslev and Wooldridge (1992).

The nonlinear filtering technique based on Taylor series can be applied in a number of applications in Finance and Economics such as stochastic volatility estimation, predictability

[1]The models nests representations such as the Fama and French (1988) and Lamoureux and Zhou (1996) permanent and temporary components; moreover, the dynamics of the logarithm of the conditional volatility nests different model specifications for the stochastic volatility models such as the proposed by Andersen and Sørensen (1996), Kim, Shephard, and Chib (1998), Jacquier, Polson, and Rossi (2004) and Jacquier, Johannes, and Polson (2007)

[2]See Judd (1998) and Schmitt-Grohe and Uribe (2004).

[3]Fernández-Villaverde and Rubio-Ramírez (2007) uncovers significant biases that arise from first and second order approximations to the functions that characterize the system.

of stock returns, fixed income instruments and structural estimation among others. My Monte Carlo experiments indicate that the Taylor series filter outperforms the unscented Kalman Filter and the extended Kalman filter. Based on different model representations, I investigate the accuracy of the filter as well as the finite sample performance of parameter estimates and find evidence that the Taylor Series filter us asymptotically unbiased. Through my experimental study I find that the filter proposed in this paper is statistically consistent, in the sense that, as the number of observations increases, the parameter estimate converges to the true parameter value.

A number of applications involve nonlinear VAR processes. These nonlinearities complicate the filtering process as well as the parameter inference procedures since the Kalman filter is not an optimal solution anymore. Consequently, different lines of research have emerged. One line of research is based on the so called deterministic filtering and uses deterministic recursions to compute the mean and variance of the state variables given the observed information. The first solution in this line is the Extended Kalman Filter (EKF) which consists of approximating the nonlinear state space with linear functions. As a result, the standard Kalman filter can be used for its estimation. Unfortunately, this approach suffers from the approximation error incurred by linearization and from the inaccuracy incurred by the fact that the posterior estimates of the states are non-normal as documented by Fernández-Villaverde and Rubio-Ramírez (2007). Julier and Uhlmann (1997) propose the Unscented Kalman Filter (UKF) to approximate the posterior density by considering a deterministically chosen set of points instead of just the conditional mean of the state. The UKF has been proven to be a powerful and efficient tool for nonlinear filtering problems in Finance and Economics. Although the UKF is accurate up to a second order for any nonlinearity, Fernández-Villaverde and Rubio-Ramírez (2007) find evidence that the inaccuracy cannot be fully solved. A recent extension of this approach is the Smolyak Kalman filter, proposed by Winschel and Krätzig (2010), which is an extension of the standard UKF based on Smolyak quadratures.

The second line of research for nonlinear filtering is based on Monte Carlo techniques. These filtering techniques are called particle filtering techniques and are based on Monte Carlo simulation with sequential importance sampling. The overall goal is to directly implement optimal Bayesian estimation by recursively approximating the complete posterior state density through simulation methods (Pitt and Shephard (1999) and Crisan and Rozovskii (2011)). The current approach to the evaluation of the likelihood of nonlinear state space models is dominated by particle filters and its extensions described in Doucet, De Freitas, and Gordon (2001). The filters with Taylor series may be an alternative approach for state estimation.

The approach proposed in this paper falls into the deterministic filtering literature and extends the current techniques for nonlinear filtering. It can be considered as a higher order efficient filtering technique that overcomes a number of difficulties that have been previously documented before for a number of reasons, first, allows arbitrary nonlinearities; second, the accuracy level is exogenous and chosen by the researcher; third, the filtering calculations are as efficient as the standard Kalman Filter; finally, the filter can be jointly used for parameter and state inference.

A closely related paper is Tanizaki and Mariano (1996), in which the Taylor series approximations in filtering are discussed as well as the biases that arise while taking the first and second order approximations; moreover, the use of Monte Carlo simulations is proposed to avoid numerical integration via Taylor series. My paper formalizes the use of Taylor series approximations of any order in deterministic filtering and proposes the use of recursive formulas for the calculation of the integrals involved in the filtering recursions. The biases that arise while taking first and second order approximations disappear as the order of approximation of the nonlinear functions increases or if the moments used for the calculation have a closed form expression.

The paper is structured as follows. Section 2 describes the general filtering problem, followed by Section 3 that describes the filtering techniques with Gaussian approximations. Section 4 describes the filtering via Taylor series and section 5 describes the Quasi-Maximum Likelihood approach for parameter estimation in filtering. In Section 6 three different applications are presented as well as the description of data, results and empirical findings. Finally, I conclude in Section 7.

2 Non-Linear Filtering

State-space models are mathematical tools that are commonly used to represent dynamic systems that involve unobserved state variables.[4] A state space representation is characterized by a state transition and a state measurement model. The state transition reflects the time evolution of the state variables, whereas the state measurement relates the unobserved state vector and the observed variables. Let x_t denote an $n-$dimensional vector that represents the state of the system at time t and y_t is a $p-$dimensional vector of observables. The states of the system follow a first order Markov process and the observations are assumed to be conditionally independent given the states. The state space model is characterized by the state transition and state measurement densities, denoted by $p(x_t|x_{t-1})$ and $p(y_t|x_t)$, respectively.

A number of applications characterize the state transition and measurement densities through

[4]See Hamilton (1994) , Kim and Nelson (1999) and Crisan and Rozovskii (2011) for more details .

the so called transition and measurement equations, which are expressed as follows

$$y_t = h(x_t) + v_t \tag{1}$$

$$x_{t+1} = g(x_t) + \varepsilon_{t+1} \tag{2}$$

where v_t and ε_t are $p-$dimensional and $n-$dimensional distributed noise vectors with variance-covariance matrices R and Q, respectively. The mappings h and g represent the measurement and deterministic process models.

To complete the specification of the model, it is assumed that the initial state of the system, x_0 has a known prior distribution, denoted by $p(x_0)$. The optimal non-linear filtering problem is to find the distribution of the state vector, x_t, given the set of observations available $y_1...y_t$. The posterior density of the states conditional on the history of observations, denoted by $p(x_0, x_1, ..., x_t|y_1, ..., y_t)$, constitutes the complete solution to the filtering problem. For tractability purposes, it is of interest to model the marginal distribution, or marginal density, denoted by $p(x_t|y_1, ..., y_t)$. If h and g are linear and v_t and ε_t are normally distributed then the solution to the filtering problem, $p(x_t|y_1, ..., y_t)$, is a Gaussian density with mean and variance that are constructed recursively according to the standard Kalman filter.[5]

3 Filtering based on Taylor Series Expansions

A number of solutions have been proposed to solve the non-linear filtering problem. The extended Kalman filter (EKF) linearizes the state space equations and then applies the Kalman (1960) filter. An improvement is the deterministic unscented Kalman filter (UKF) proposed by Julier and Uhlmann (1997).

The UKF relies on the idea that approximating the moments of a transformed random variables is simpler than approximating its density function. The unscented filter approximates the first two moments needed for the Kalman update. The approximation is based on quadrature techniques where the number of grid points is taken to be $2d + 1$, where d is the dimension of the integrand. As Winschel and Krätzig (2010) emphasizes, the UKF is an attempt to solve the curse of dimensionality; however, the filter generates another curse in terms of approximation errors. As the UKF raises the number of points only linearly, the effect is that accordingly the accuracy of the numerical integration decreases with the dimensionality and non-linearity of the integrands. Therefore, the UKF's error of the likelihood approximation comes from restricting the approximation to two moments and their ad hoc approximation. The unscented filter, is

[5] See Kalman (1960).

therefore, restricted to a low polynomial exactness and a small number of states.

The filter based on Taylor series expansions avoids the ad-hoc moment approximation and instead uses exact moment calculations or a level of approximation exogenously chosen by the researcher. The moments are then updated in the usual way by the Kalman gain in the filtering step. An advantage of this procedure compared to other deterministic filters is that the approximation level can be chosen according to the problem at hand and that the filter is also useful for other than normally distributed shocks. Finally, the computational efficiency of the filter is comparable with standard deterministic filters such as the EKF or UKF.

The notation $\mathcal{N}(z;\mu,\Sigma)$ is a shorthand for the density of a multivariate normal distribution with argument z, mean μ, and covariance Σ. As in the UKF and EKF, I assume that the initial state density is normal with mean x_0 and covariance matrix P_0. I also assume that the densities involved in the filtering steps are Gaussian. In this case, the previous-posterior density is

$$p\left(x_t\left|y_{1,t}\right.\right)=\mathcal{N}\left(x_t;x_{t|t},P_{t|t}\right);$$

the prior density

$$p\left(x_{t+1}\left|y_{1:t}\right.\right)=\mathcal{N}\left(x_{t+1};x_{t+1|t},P_{t+1|t}\right)$$

is characterized by the first two moments of Equation (2), which are given by

$$x_{t+1|t} = \mathbb{E}\left[g\left(x_t\right)\left|y_{1,t}\right.\right], \tag{3}$$

$$P_{t+1|t} = Var\left[g\left(x_t\right)\left|y_{1,t}\right.\right]+Q. \tag{4}$$

Similarly, the measurement density, defined by the observation equation in 1, is Gaussian,

$$p\left(y_{t+1}\left|y_{1:t}\right.\right)=\mathcal{N}\left(y_t;y_{t+1|t},P^{yy}_{t+1|t}\right),$$

with mean

$$y_{t+1|t}=\mathbb{E}\left[h\left(x_{t+1}\right)\left|y_{1,t}\right.\right] \tag{5}$$

and variance-covariance matrix

$$P^{yy}_{t+1|t}=Var\left[h\left(x_{t+1}\right)\left|y_{1,t}\right.\right]+R, \tag{6}$$

where R is the covariance matrix of the measurement shocks. Moreover, the covariance between the observed and unobserved variables is represented by

$$P^{xy}_{t+1|t}=Cov\left[x_{t+1},h\left(x_{t+1}\right)\left|y_{1,t}\right.\right]. \tag{7}$$

The recursion is closed by the filtering step and the posterior density is obtained according to the usual Kalman update, represented by the following set of recursive equations[6]:

$$\begin{aligned} p\left(x_{t+1}\left|y_{1:t+1}\right.\right) &= \mathcal{N}\left(x_{t+1}; x_{t+1|t+1}, P_{t+1|t+1}\right), \\ K_{t+1} &= P^{xy}_{t+1|t}\left(P^{yy}_{t+1|t}\right)^{-1}, \\ x_{t+1|t+1} &= x_{t+1|t} + K_{t+1}\left(y_{t+1} - y_{t+1|t}\right), \\ P_{t+1|t+1} &= P_{t+1|t} - K_{t+1}P^{yy}_{t+1|t}K_{t+1}. \end{aligned} \tag{8}$$

The previous filtering recursions rely on the calculations of the moments in Eqs. (3) - (7), which are expected values of potentially non-linear transformations of random variables which may not have a closed form solution. If the moments have a closed form expression, then the filtering recursions are as computationally efficient as the standard Kalman filter, since only function evaluations are required. For a number of problems, the moments involved in the Kalman filter recursions have a closed form [7]; however, numerical approximations are required if they do not.

A natural approach consists of approximating the observation and transition equations with Taylor series expansions; consequently, the moments of the observation and transition equations involved are calculated with the expected values of its Taylor approximations. The numerical integration problem is solved by calculating the derivatives of the observation and transition equations as well as the joint moments of normally distributed random vectors. The variance-covariance matrices as well as the moments of the normally distributed vectors are calculated based on a set of efficient recursions recently proposed in Savits (2006) that have been applied in Finance by Garlappi and Skoulakis (2010) and Garlappi and Skoulakis (2011). One advantage of this procedure is that the order of approximation is exogenous and can be arbitrarily high.

A detailed description of the calculation of moments via Taylor series approximations is provided in the following section.

4 Taylor Series Approximations

Let $y = f(x)$ denote a smooth function, where $f : \mathbb{R}^N \longrightarrow \mathbb{R}$, and let μ denote an $N-$dimensional constant vector.[8] Let $\mathbf{q} = (q_1, ..., q_N)$ denote a vector of nonnegative integers, $|\mathbf{q}| = \sum_{n=1}^{N} q_n$,

[6] For a detailed description of the standard Kalman filter, please see the appendix.

[7] The analytical tractability and efficiency of closed form expressions has not been used extensively in the filtering literature; an exception is Ito and Xiong (2000) and proposes the use of Gaussian densities for non-linear filtering.

[8] I will follow the convenient tensor notation from Savits (2006) and Garlappi and Skoulakis (2010).

$\mathbf{q}! = \prod_{n=1}^{N} (q_n!)$, and $f_{\mathbf{q}}(\mu)$ denote the partial derivative of order $\mathbf{q}$ of the function $f(\mu)$ evaluated at μ i.e.,

$$f_{\mathbf{q}}(\mu) = \frac{\partial^{q_1+...+q_N} f}{\partial x_1^{q_1} ... \partial x_N^{q_N}} (\mu) . \tag{9}$$

The following theorem states preamble of the Taylor series approximations

Theorem 4.1 *Let $U \subset \mathbb{R}^N$ be an open subset, $x \in U$, $\mu \in \mathbb{R}^N$, so that $tx + (1-t)\mu \in U$ for all $t \in [0,1]$. Assume $f : U \longrightarrow \mathbb{R}$ is $(M+1)-$times continuously differentiable. Then, there is a $\lambda \in [0,1]$, so that*

$$f(x) = \sum_{\{\mathbf{q}:|\mathbf{q}|\leq M\}} \frac{1}{\mathbf{q}!} f_{\mathbf{q}}(\mu) \prod_{n=1}^{N} (x_n - \mu_n)^{q_n} + \sum_{|\mathbf{q}|=M+1} \frac{1}{\mathbf{q}!} f_{\mathbf{q}}(\zeta) \prod_{n=1}^{N} (x_n - \mu_n)^{q_n} ,$$

where $\zeta = \lambda x + (1-\lambda)\mu$.

Remark An immediate corollary of Theorem 4.1 is the generic M-th order Taylor approximation of f centered at μ, which is defined as

$$f(x) \simeq \sum_{\{\mathbf{q}:|\mathbf{q}|\leq M\}} \frac{1}{\mathbf{q}!} f_{\mathbf{q}}(\mu) \prod_{n=1}^{N} (x_n - \mu_n)^{q_n} , \tag{10}$$

Now, suppose that $x \sim N(\mu, \Sigma)$ and we are interested in calculating the expected value of $f(x)$. From Eq. (10) we can approximate the expected value as

$$\mathbb{E}[f(x)] \simeq \sum_{\{\mathbf{q}:|\mathbf{q}|\leq M\}} \frac{1}{\mathbf{q}!} f_{\mathbf{q}}(\mu) \mathbb{E}\left[\prod_{n=1}^{N} (x_n - \mu_n)^{q_n}\right] . \tag{11}$$

Intuitively, Equation (11) provides an approximation for the expected value of a transformation of a normally distributed random vector which is based on two separate elements: the $M-$ order derivatives of the function f and the cross moments of a normally distributed random vector. In most of the applications, the derivatives of the function f have an analytical expression and can be calculated explicitly. As for the moment calculation, Savits (2006) provides efficient recursions based on Faà di Bruno's formula to compute the joint moments of a normally distributed random vector. This recursive formulation is summarized in Proposition 4.2.

4.1 Efficient computation of the moments of a multivariate normal distribution

Let $X = (x_1, x_2, ..., x_N)$ denote a multivariate normal random vector with zero-mean vector and covariance matrix Σ, where the component i, j denotes the covariance between the random variables x_i and x_j equal to σ_{ij}. Let $\mu_{(q_1,...,q_N)}$ be its $(q_1, ..., q_N)$- moment, where $q_1, ..., q_N$ are nonnegative integers, i.e. $\mu_{(q_1,...,q_N)} = \mathbb{E}[x_1^{q_1}...x_N^{q_N}]$. Then, from Theorem 5.1 in Savits (2006), we have the following recursive relation between the multivariate moments of X.

Proposition 4.2 *Set $\mu_{(0,...,0)} = 1$. Then, for all $(q_1, ..., q_N) \geq \mathbf{0}_N$ and $1 \leq j \leq N$, we have*

$$\mu_{(q_1,...,q_N)+e_j} \equiv \mathbb{E}\left[x_1^{q_1}...x_j^{q_j+1}...x_N^{q_N}\right] = \sum_{k=1}^{N} \sigma_{jk} q_k \mu_{(q_1,...,q_N)-e_k}, \tag{12}$$

where e_j is the N-dimensional unit vector with $j-th$ component equal to 1 and all the other components equal to zero.

Proof See Savits (2006).

4.2 Efficient Calculation of Variance-Covariance matrices

To calculate the variances and covariances involved in the Kalman filter recursions, I approximate the square and the product of a transformed random variable with its Taylor series around its mean. [9] Based on these approximations, all the expected values involved in the computation of variances and covariances are calculated.

The variance of any random variable requires the calculation of its first two moments. Following the same intuition as Equation (11), the second moment of f, $\mathbb{E}\left[f^2\right]$ is calculated based on its Taylor Series expansions, i.e.

$$\mathbb{E}\left[f^2(x)\right] \approx \sum_{\{\mathbf{q}:|\mathbf{q}|\leq M\}} \frac{1}{\mathbf{q}!} \left(f^2\right)_{\mathbf{q}}(\mu)\, \mathbb{E}\left[\prod_{n=1}^{N}(x_n - \mu_n)^{q_n}\right] \tag{13}$$

where $\left(f^2\right)_{\mathbf{q}}(\mu)$ denotes the partial derivative of order $\mathbf{q}$ of the square of the function f evaluated at μ.

Calculating the derivatives of f^2 may be quite cumbersome since it involves the derivative of the composition of two functions; however, Proposition 4.3 provides a general recursive algorithm to calculate the derivative of the square of a function based on the Faà di Bruno formula for the derivative of the composition of two functions. These expressions will be used to calculate the variances in Eqs. (4) and (6) used in the Kalman Filter update.

[9] The choice of the mean vector, μ, as a center of expansion of the Taylor series is convenient for two reasons: first, all the calculations that involve derivatives are independent of the expectation operator; and second, $\mathbb{E}[\prod_{n=1}^{N}(x_n - \mu_n)^{q_n}] = 0$, for all vectors $\mathbf{q}$ such that $\sum_{n=1}^{N} q_n$ is an odd number. In any case, the results will still valid if the chosen center of expansion is any other constant vector. If we are dealing with conditional expectations, the results will still be valid; moreover, any measurable random vector can be chosen as a center of expansion.

Proposition 4.3 *Let $f : \mathbb{R}^N \longrightarrow \mathbb{R}$ be an $(M+1)-$ times continuously differentiable function, then the derivatives of $g(x) = f(x)^2$ can be obtained from the following vector recursion*

$$\begin{aligned} g_{\mathbf{0}}(\mathbf{x}) &= f(\mathbf{x})^2 \\ g_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x}) &= \sum_{\left\{\ell \in \mathbb{N}_0^N : \mathbf{0}_N \leq \ell \leq \mathbf{q}\right\}} 2 \times \binom{\mathbf{q}}{\ell} f_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x}) f_\ell(\mathbf{x}) \end{aligned} \tag{14}$$

Proof *See the Appendix.*

As a result, the variance is computed as the difference between the second moment of the function and the squared approximation of the first moment, that is

$$Var[f(x)] = \mathbb{E}\left[f^2(x)\right] - \mathbb{E}^2[f(x)].$$

A counterpart for Proposition 4.3 that allows for the calculation of covariances is given by Proposition 4.4; in which the covariances involved in the filtering recursions are calculated via Taylor series of the product of two functions. As in the variance, the idea is to approximate the expected value of the product of two functions using the derivatives evaluated in the vector of means and the joint moments of a normally distributed random vector, as in Eq. (13), that is

$$\mathbb{E}[f_1(x) f_2(x)] \approx \sum_{\{\mathbf{q}:|\mathbf{q}|\leq M\}} \frac{1}{\mathbf{q}!} (f_1 \cdot f_2)_{\mathbf{q}}(\mu) \mathbb{E}\left[\prod_{n=1}^{N} (x_n - \mu_n)^{q_n}\right]. \tag{15}$$

where $(f_1 \cdot f_2)_{\mathbf{q}}(\mu)$ denotes the partial derivative of order $\mathbf{q}$ of the product of the functions f_1 and f_2 evaluated at the constant point μ.

Proposition 4.4 *Let $f_1, f_2 : \mathbb{R}^N \longrightarrow \mathbb{R}$ be $(M+1)-$ times continuously differentiable functions. Let $g(\mathbf{x}) = f_1(\mathbf{x}) f_2(\mathbf{x})$,then the derivatives of $g(x)$, are given by*

$$\begin{aligned} g_{\mathbf{0}}(\mathbf{x}) &= f_1(\mathbf{x}) f_2(\mathbf{x}) \\ g_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x}) &= \sum_{\left\{\ell \in \mathbb{N}_0^N :\ \mathbf{0}_N \leq \ell \leq \mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} f_{1,\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x}) f_{2,\ell}(\mathbf{x}), \\ &\quad + \sum_{\left\{\ell \in \mathbb{N}_0^N :\ \mathbf{0}_N \leq \ell \leq \mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} f_{2,\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x}) f_{1,\ell}(\mathbf{x}). \end{aligned} \tag{16}$$

Proof *See the Appendix.*

Now, based the covariance between the two transformations is calculated as

$$cov[f_1(x), f_2(x)] = \mathbb{E}[f_1(x) f_2(x)] - \mathbb{E}[f_1(x)] \mathbb{E}[f_2(x)].$$

Finally, the covariance matrix involved in the Kalman filter step in Equation (7), can be obtained through Proposition 4.4. However, a convenient shortcut is obtained via Stein's Lemma, the only requirement being the integrability of the derivative of a function f.

Lemma 4.5 *(Stein's Lemma)For any function $f(x_1, ..., x_N)$ such that $\partial f / \partial x_i$ exists almost everywhere and $\mathbb{E}\left|\frac{\partial}{\partial x_i} f(X)\right| < \infty$, $i = 1, ..., n$. Let $\nabla f(X) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)^T$. Then the following identity holds*

$$cov(X, f(X)) = \Sigma \times \mathbb{E}[\nabla f(X)], \tag{17}$$

more specifically

$$cov(x_1, f(x_1, ..., x_N)) = \sum_{i=1}^{N} cov(x_1, x_i) \times \mathbb{E}\left[\frac{\partial}{\partial x_i} f(x_1, ..., x_N)\right]. \tag{18}$$

Proof See the Appendix.

Recall that $\nabla f(x) = \left[\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_N}\right]^\top$, represents the vector of partial derivatives of the function f. In this case, if the function f is $(M+1)-$times continuously differentiable, then the power series of the $i-th$ component can be calculated directly and is represented by

$$\frac{\partial f}{\partial x_i} \simeq \sum_{\{\mathbf{q}:0<|\mathbf{q}|\leq M\}} \frac{1}{\mathbf{q}!} f_{\mathbf{q}}(\mu) \times q_i (x_i - \mu_i)^{q_i - 1} \times \prod_{\substack{n=1 \\ n\neq i}}^{N} (x_n - \mu_n)^{q_n}.$$

5 Quasi-Maximum Likelihood Parameter Estimation

The filters based on Gaussian approximations have been applied recently for state estimation as well as for inference purposes. [10] In this section, I introduce a quasi-maximum likelihood method for parameter estimation of nonlinear state space representations based on Bollerslev and Wooldridge (1992).

5.1 Quasi-Likelihood Function

For each time observation, a conditional mean, $y_{t+1|t}$, and conditional covariance, $P^{yy}_{t+1|t}$, is calculated through the Extended, Unscented or Higher order Taylor series approach. A quasi log-likelihood function is constructed assuming that the observation, y_{t+1}, is normally distributed

[10] See Christoffersen, Jacobs, Karoui, and Mimouni (2012), van Binsbergen and Koijen (2011), Campbell, Sunderam, and Viceira (2011) and Calvet, Fisher, and Wu (2010) have used the Unscented Kalman Filter in different applications in Financial Economics.

with mean, $y_{t+1|t}$, and volatility, $P^{yy}_{t+1|t}$. Let θ denote the vector of parameters that are used to perform the Kalman Filter. The log-likelihood for each observation, denoted by $l_t(\theta)$, is calculated as

$$l_t(\theta) = -\frac{p}{2}\log(2\pi) - \frac{1}{2}\log\left(P^{yy}_{t+1|t}\right) - \frac{1}{2}\frac{\left(y_{t+1} - y_{t+1|t}\right)^2}{P^{yy}_{t+1|t}}. \tag{19}$$

Finally, we choose the parameter values θ, that maximize

$$L(\theta) = \Sigma_{t=1}^{T} l_t(\theta). \tag{20}$$

Gallant and White (1988) develop a statistical theory for misspecified models, known as Quasi Maximum Likelihood (QML) that can be applied to guarantee that the QML function, (19), is well defined. Moreover, Bollerslev and Wooldridge (1992) show that a Gaussian QML estimator is consistent, under a set of regularity conditions. More specifically, Bollerslev and Wooldridge (1992) show that the true but unknown vector of parameters is the global maximizer of (20) if the conditions following conditions hold:

$$\begin{aligned} \mathbb{E}\left[v_{t+1} \,|y_{1:t}\right] &= 0 \\ Var\left[v_{t+1} \,|y_{1:t}\right] &= R. \end{aligned}$$

As for the calculation of asymptotic standard errors, Gallant and White (1988) show that under certain regularity conditions, the covariance matrix of the QML estimator has a closed form expression. [11]

6 Applications

In this section, I test and compare the performance of my nonlinear filtering methodology to that of the Extended Kalman Filter, the Unscented Kalman filter and the Gaussian Kalman filter for three different model specifications that have been proposed in Financial Economics. The first application is the Stochastic Volatility Model by Andersen and Sørensen (1996), Andersen, Bollerslev, Diebold, and Ebens (2001), Andersen, Bollerslev, Diebold, and Labys (2003) and Broto and Ruiz (2004). The second application is a return representation that was firstly proposed by Brandt and Kang (2004) and recently used by Boguth, Carlson, Fisher, and Simutin (2011) in the conditional asset pricing literature. The third representation is a simple version of the Dynamic Stochastic General Equilibrium model studied by Schmitt-Grohe and Uribe (2004) and Flury and Shephard (2011).

[11] See the Appendix for details.

6.1 Stochastic Volatility Models

The standard stationary stochastic volatility[12] model in discrete time is represented by

$$\begin{aligned} y_t &= \eta_t \cdot \sigma_t \\ \log \sigma_t^2 &= d + \phi \log \sigma_{t-1}^2 + \varepsilon_t,\ \varepsilon_t \sim N\left(0, \widetilde{\sigma}_\varepsilon^2\right), \end{aligned} \tag{21}$$

where y_t is the return observed at time t and σ_t is the corresponding volatility; d is a scale parameter for the volatility process, η_t is a white noise process with unit variance that represents the innovations in the level or returns. The disturbance of the volatility equation, ε_t is assumed to be a Gaussian white noise process; finally $|\phi|$ is considered as a measure of persistence of shocks to the volatility. The variance of the log-volatility process, $\widetilde{\sigma}_\varepsilon^2$, measures the uncertainty of future volatility. The log normality specification for the volatility is consistent with Andersen, Bollerslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, Diebold, and Labys (2003) that show that the log-volatility process can be well approximated by a Normal distribution and with Taylor (2008) who proposes to model the logarithm of volatility as an AR(1) process. When ϕ is close to one and $\widetilde{\sigma}_\varepsilon^2$ is close to zero then the evolution of volatility over time is very smooth; however, in the limit, if $\phi = 1$ and $\widetilde{\sigma}_\varepsilon^2 = 0$, the volatility is constant over time, and consequently, the returns are homoscedastic. As noted by Broto and Ruiz (2004), if $\widetilde{\sigma}_\varepsilon^2 = 0$ the model cannot be identified.

6.1.1 State Space Representation and Implementation

An alternative representation of Equation (21) can be obtained by re-parameterizing the stochastic volatility process as

$$\begin{aligned} y_t &= \overline{\sigma} \exp(s_t)\, \eta_t \\ s_t &= \phi s_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right). \end{aligned} \tag{22}$$

Clearly, the state variables are $x_t = [s_t, \eta_t]^\top$, $h(x_t) = h(s_t, \eta_t) = \overline{\sigma} \exp(s_t)\, \eta_t$, the random noise of the observation equation, v_t, is identically zero and as a result its variance is identically zero ($R \equiv 0$).[13] Furthermore, we can treat η_t as a state variable that only depends on the current shock and does not depend on its lagged values; as a result we have a more flexible framework that potentially handles correlated shocks between the observation and transition equations; in

[12] See Ghysels, Harvey, and Renault (1996) and Shephard (2005) for a comprehensive review.

[13] This is a unique feature of the Gaussian filters; most of the simulation based filters, such as the Particle filter, require that all the variances of the transition and measurement equations should be positive semi-definite.

other words,

$$
\begin{aligned}
\begin{bmatrix} s_t \\ \eta_t \end{bmatrix} &= \begin{bmatrix} \phi & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} s_{t-1} \\ \eta_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \quad \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim N(0, Q) \\
Q &= \begin{bmatrix} \sigma_\varepsilon^2 & \rho \\ \rho & 1 \end{bmatrix},
\end{aligned}
$$

where $\rho \equiv corr(s_t, \eta_t)$. Two transition equations are necessary for this representation; the first equation is given in (22) while the second transition equation characterizes the random shock that characterizes the observation equation. Although all the moments involved the filtering algorithms exist and have a closed form expression, I approximate the expected values and covariances involved in the filtering calculations with Taylor series to analyze the precision of the method for state and parameter estimation. [14]

According to Section 2, we need to approximate the first two moments of the returns with stochastic volatility, that is

$$
y_{t+1|t} = \mathbb{E}\left[y_{t+1} \,|Y_t\right] = \mathbb{E}\left[\overline{\sigma} \exp\left(s_{t+1}\right) \eta_{t+1} \,|Y_t\right]
$$

and

$$
\begin{aligned}
P^{yy}_{t+1|t} &= var\left[y_{t+1} \,|Y_t\right] = \mathbb{E}\left[\overline{\sigma}^2 \exp\left(2s_{t+1}\right) \eta^2_{t+1} \,|Y_t\right] - \mathbb{E}^2\left[\overline{\sigma} \exp\left(s_{t+1}\right) \eta_{t+1} \,|Y_t\right] \\
&= \mathbb{E}\left[\overline{\sigma}^2 \exp\left(2s_{t+1}\right) \eta^2_{t+1} \,|Y_t\right] - y^2_{t+1|t}.
\end{aligned}
$$

Finally, the computation of the Kalman gain involves calculation of the covariances between the observation and transition equations, that is

$$
P^{xy}_{t+1|t} = cov \begin{bmatrix} s_{t+1}, y_{t+1} \,|Y_t \\ \eta_{t+1}, y_{t+1} \,|Y_t \end{bmatrix} = P_{t+1|t} \times \mathbb{E} \begin{bmatrix} \overline{\sigma} \exp\left(s_{t+1}\right) \eta_{t+1} \,|Y_t \\ \overline{\sigma} \exp\left(s_{t+1}\right) |Y_t \end{bmatrix},
$$

where the last equality comes from applying the multivariate version of Stein's lemma shown in Section 4.[15]

[14] See Broto and Ruiz (2004) for an extensive review of stochastic volatility model estimation.

[15] The Appendix contains closed form expressions of the expected values, variances and covariances involved in the Kalman Filter recursions of the Stochastic Volatility Model.

6.1.2 Monte Carlo Simulation Results

In this section I conduct a Monte Carlo study to test for the accuracy of the filters with Taylor approximations in two different ways. The first exercise consists testing the accuracy of the state estimates and is explained as follows. A time series of 1000 observations is generated using the model in (22), assuming $\phi = 0.98$, $\sigma_\varepsilon = 0.1414$, $\overline{\sigma} = 1$ and $\rho = -0.5$.[16] Conditional on the simulated noisy observations, y_t, the log-volatility processes, s_t, was filtered using different orders of approximation, M. For simplicity I will denote by TKF-M the filter with an $M-th$ order Taylor series approximation.

The experiment was repeated 250 times with a random re-initialization for each run. The EKF and UKF were included for comparison purposes. The UKF parameters were set to $\alpha = 1, \beta = 0$ and $\kappa = 2$. These parameters are optimal for the scalar case. Finally, the parameter estimates of the filter that uses closed form expressions of the conditional moments is included for comparison purposes. I will refer to this filter as the Gaussian Filter (GF hereafter).[17].

[Insert Table 1 about here]

Table 1 summarizes the performance of the different filters. The table shows the means and standard deviations of the mean-square-error (MSE) of the state estimates. The first two columns contain statistics of the mean squared errors (MSE) of the simulated log-volatility versus its filtered estimate. The minimum MSE is achieved by the filter with fourth and fifth order of approximation. In this case, the filters with 10-th and 11-th order of approximation and the Gaussian filters provide the same MSE statistics. The MSE for two consecutive orders of approximation (even and odd) are exactly the same since the joint moments of an odd order of a normally distributed random vector are equal to zero. Although the Unscented Kalman Filter is commonly known as a second order filter, my simulation results show evidence that the MSE results of the second order filter are slightly different from the statistics obtained from the Unscented Kalman Filter.

[Insert Figure 1 about here]

[16] These parameter values have been extensively used in Broto and Ruiz (2004) as well as in empirical applications of daily returns.

[17] Ito and Xiong (2000) propose the name of Gaussian filters to the filtering methods that use Gaussian densities to approximate the posterior densities of the filtering problem.

Figure 1 compares the state estimates of the log-volatility process generated from a single run using different filters. The tests were conducted using all the orders of approximation, but only reported the fifth order for easiness of exposition. The first panel compares the simulated series with the filtered estimates based on the fifth order approximation and the Gaussian filters; the difference between both filtered series is almost negligible. The second panel contains the simulated series with the the filtered series based on the EKF and UKF. For the first 400 observations, all the filters perform similarly; however, for the remaining set of observations, the UKF and EKF provide more volatile state estimates compared to the other filters. In general, the UKF provides more volatile state estimates than higher order filters.

[Insert Figure 2 about here]

As with most nonlinear models, it is difficult , if not impossible, to prove that the parameters of the model are uniquely identified. In order to analyze the uniqueness of the QMLE estimates, I implemented the following procedure. For a set of parameter values, a path of noisy returns was simulated and a quasi-likelihood function was constructed based on the simulated path. An initial identification exercise was performed by calculating the quasi-likelihood function in the set of parameters used for the simulation and ranging independently each parameter ϕ, σ_{ε}, $\overline{\sigma}$ and ρ. The results are shown in Figure 2; the dashed lines represent the parameter values that were used to simulate the data. The concavity of the quasi-likelihood function with respect to each parameter shows evidence that all the parameters are well identified and the maxima are achieved in parameter values close enough to the ones used to simulate the data.

An alternative way to analyze the finite sample properties of the QML estimator is via Monte Carlo simulation. In particular, I repeatedly estimated the model from 250 independent samples of $T = 500$ monthly returns with the parameter values used in the previous section via QML estimation. For different orders of approximation M, the QML function based on the Taylor series approximation as well as the UKF. Table **??** presents the results. It shows the true parameter values in the first row and describes the sample mean and variance of the corresponding parameter estimates in the remaining rows. The average estimates under the different filtering algorithms are all close to the true parameter values, suggesting that the estimators via Taylor series are relatively unbiased. The estimates for $\overline{\sigma}$ under the EKF, UKF and the third order approximations are biased. Moreover, the standard deviations of the QML estimates of the 9-th approximations are similar to the standard errors of the third order approximation. This effect may be caused by a small sample bias as well as the effect of numerical errors.

[Insert Table ?? about here]

6.1.3 Consumption Growth

A standard model for the log of consumption follows a random walk with drift μ_c and innovation standard deviation, $\overline{\sigma}$,

$$\begin{aligned} \Delta \ln(C_{t+1}) &= \mu_c + \overline{\sigma}\eta_{t+1} \\ \eta_{t+1} &\sim N(0,1). \end{aligned}$$

The previous specification has been analyzed in Tallarini (2000) and Barillas, Hansen, and Sargent (2009). However, there is evidence of time variation in the conditional deviation of may macroeconomic series, as documented in Stock and Watson (2002), McConnell and Perez-Quiros (2000), Fernández-Villaverde and Rubio-Ramírez (2007), Justiniano and Primiceri (2008), Clark (2009) and Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2012), among others.

As a result, Bidder and Smith (2011) propose an alternate endowment process that features stochastic volatility in log-consumption growth, that is the standard stochastic volatility model, as in (22), that is represented as follows:

$$\begin{aligned} \Delta \ln(C_{t+1}) - \mu_c &= \overline{\sigma} \exp(s_{t+1})\, \eta_{t+1},\ \eta_t \sim N(0,1) \\ s_{t+1} &= \phi s_t + \varepsilon_{t+1},\ \varepsilon_{t+1} \sim N\left(0, \sigma_\varepsilon^2\right). \end{aligned} \tag{23}$$

For parameter estimation purposes, I use the monthly vintage series from the Federal Reserve Bank of Philadelphia to construct the real consumption per capita from January 1959 to March 2012. The "monthly" log-consumption growth data was constructed using the real time data of real personal consumption expenditures in nondurables and services from the Real-Time Data Set for Macroeconomists from the Federal Reserve Bank of Philadelphia.[18] This "real-time data set" of macro economic variables was created to update and verify the accuracy of forecasting models of macro variables and provides snapshots of the macroeconomic data available at any given date in the past. [19]

The summary statistics and time-series are shown in Table 3 and Figure 3.

[Insert Table 3 about here]

[Insert Figure 3 about here]

[18]http://www.phil.frb.org/research-and-data/real-time-center/real-time-data

[19]See Croushore and Stark (2001) for details.

As in the previous sections, the parameter values were estimated using the quasi-likelihood function constructed based on the Kalman filter with Taylor series approximations of order $M = 1, 3, 5, 7, 9$ as well as the gaussian filters. The estimates using the UKF were included, for comparison purposes. The results are shown in Table 4. In this case, all the parameters were identified. The parameter estimates using the QML function constructed with an order of 3 or more have similar values; however, the adjusted standard errors change as the order of approximation changes. This is due to the fact that the quasi-likelihood function achieves the same value starting from an order of 5 or more. In general, the magnitude of the standard errors decreases as the order of approximation increases. Overall, the parameter estimates have similar values to those found in Bansal and Yaron (2004), Bidder and Smith (2011) and Ludvigson (2012) with a slightly lower growth rate and higher variance, most likely due to the longer data series including the recession starting in the last quarter of 2007.

[Insert Table 4 about here]

6.2 Risk-Return Trade-off

Brandt and Kang (2004) introduces a non-linear representation for the return dynamics that allows for positive risk premium in the context of a latent vector autoregressive (VAR) process and is presented as follows.

Let y_t be the continuously compounded excess returns with time-series dynamics represented by

$$y_t = \mu_{t-1} + \sigma_{t-1}\varepsilon_t \text{ with } \varepsilon_t \sim N(0,1) \tag{24}$$

where $\mu_{t-1}$and σ_{t-1} represent the conditional volatility of the excess returns. In addition, it is assumed that the conditional mean and volatility are unobservable and that they follow a first order VAR process in logs:

$$\begin{bmatrix} \ln \mu_t \\ \ln \sigma_t \end{bmatrix} = d + A \begin{bmatrix} \ln \mu_{t-1} \\ \ln \sigma_{t-1} \end{bmatrix} + \eta_t \text{ with } \eta_t \sim N(0, \Sigma), \tag{25}$$

where

$$
\begin{aligned}
d &= \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}, \ A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ and} \\
\Sigma &= \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \text{ with } b_{12} = b_{21} = \rho\sqrt{b_{11}b_{22}}. \qquad (26)
\end{aligned}
$$

The first equation of the VAR in Eq. (25) describes the dynamics of the logarithm of the conditional mean, and it captures the Fama and French (1988) and Lamoureux and Zhou (1996) permanent and temporary components, in which the stock prices are governed by a random walk and a stationary random process, respectively. The second equation of the VAR describes the dynamics of the logarithm of the conditional volatility; and nests different specifications such as the proposed by Andersen and Sørensen (1996), Kim, Shephard, and Chib (1998), Jacquier, Polson, and Rossi (2004) and Jacquier, Johannes, and Polson (2007). This latent VAR approach allows to study the contemporaneous and intertemporal relationships between expected returns and risk without relying on predictors; moreover, the contemporaneous relationship between the conditional mean and volatility can be analyzed.

Following Hamilton (1994), if the VAR is stationary, the unconditional moments for the mean and volatility are given by

$$
E\begin{bmatrix} \ln \mu_t \\ \ln \sigma_t \end{bmatrix} = (I - A)^{-1}\, d \qquad (27)
$$

and

$$
\operatorname{vec}\left(\operatorname{cov}\begin{bmatrix} \mu_t \\ \sigma_t \end{bmatrix}\right) = (I - (A \otimes A))^{-1} \operatorname{vec}(\Sigma) \qquad (28)
$$

where $\otimes$ represents the Kronecker product.

The return dynamics presented in Eq. (25) has key elements: the transition matrix A and the correlation coefficient ρ. The diagonal elements of A capture the persistence of the conditional moments, and the off diagonal elements reflect the intertemporal feedback between the conditional volatility and the conditional mean. A general correlation structure is specified; the conditional mean and volatility are correlated with the return innovations; $Corr\,[\varepsilon_t, \eta_t] = [\rho_\mu, \rho_\sigma]'$.

6.2.1 State Space Representation and Implementation

The representation given in Eqs. (24) and (25) defines a state-space model; where the first equation represents a non-linear measurement or observation equation and the second equation is the standard linear transition equation. In order to infer about both the parameters of the VAR and the realizations of the conditional moments from the observed returns, we need to solve a sequence of filtering problems. Filtering generates the one step ahead forecasts of the latent variables $\mathbb{E}\left[\ln \mu_t, \ln \sigma_t \,|y_1, ..., y_t\right]$ and the corresponding forecast variances $Var\left[\ln \mu_t, \ln \sigma_t \,|y_1, ..., y_t\right]$, which in a linear Gaussian state-space model are used to construct the likelihood function. The nonlinearity of the observation equation makes the problem not standard.

A simpler representation of the state space representation can be obtained by redefining in state variables in demeaned terms, that is $m_t = \ln \mu_t - \mathbb{E}\left[\ln \mu_t\right]$ and $v_t = \ln \sigma_t - \mathbb{E}\left[\ln \sigma_t\right]$, so that $\mu_t = \overline{\mu} \exp\left(m_t\right)$ and $\sigma_t = \overline{\sigma} \exp\left(v_t\right)$, where $\overline{\mu} = \exp\left(\mathbb{E}\left[\ln \mu_t\right]\right)$ and $\overline{\sigma} = \exp\left(\mathbb{E}\left[\ln \sigma_t\right]\right)$. Finally, let $x_t = \left[x_{1t}, x_{2t}, x_{3t}, x_{4t}, x_{5t}\right]' = \left[m_{t-1}, v_{t-1}, \varepsilon_t, m_t, v_t\right]'$, then equations (24) and (25) can be rewritten as:

$$y_t = \overline{\mu} \exp\left(x_{1t}\right) + \overline{\sigma} \exp\left(x_{2t}\right) x_{3t}, \tag{29}$$

and

$$x_t = \widetilde{A} x_{t-1} + \underline{w_t} \text{ with } \underline{w_t} \sim N\left(0, \widetilde{\Sigma}\right), \tag{30}$$

where

$$\widetilde{A} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{11} & a_{12} \\ 0 & 0 & 0 & a_{21} & a_{22} \end{bmatrix},$$

and

$$\widetilde{\Sigma} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \rho_\mu \sqrt{b_{11}} & \rho_\sigma \sqrt{b_{22}} \\ 0 & 0 & \rho_\mu \sqrt{b_{11}} & b_{11} & \rho \sqrt{b_{11} b_{22}} \\ 0 & 0 & \rho_\sigma \sqrt{b_{22}} & \rho \sqrt{b_{11} b_{22}} & b_{22} \end{bmatrix}.$$

where $Corr\left[\varepsilon_t, \eta_t\right] = \left[\rho_\mu, \rho_\sigma\right]'$. All the parameters can be represented into the vector $\psi = \left[a_{11}, a_{12}, a_{21}, a_{22}, b_{11}, b_{22}, \rho, \overline{\mu}, \overline{\sigma}, \rho_\mu, \rho_\sigma\right]'$ and the matrix of conditional means and volatilities can be stacked in the following matrix:

$$\theta = \left[x_0, x_1, ..., x_{T-1}\right]'.$$

According to Section 2, the filtering and estimation problems can be solved via the filtering method with Taylor series. The first and second moments to be approximated are

$$y_{t+1|t} = \mathbb{E}\left[y_{t+1}\left|Y_t\right.\right] = \overline{\mu}\mathbb{E}\left[\exp\left(x_{1t+1}\right)\left|Y_t\right.\right] + \overline{\sigma}\mathbb{E}\left[\exp\left(x_{2t+1}\right)x_{3t+1}\left|Y_t\right.\right],$$

and

$$P^{yy}_{t+1|t} = Var\left[y_{t+1}\left|Y_t\right.\right] = \mathbb{E}\left[y^2_{t+1}\left|Y_t\right.\right] - y^2_{t+1|t}.$$

Finally, the covariance term involved in the Kalman gain, is calculated as

$$P^{xy}_{t+1|t} = cov\begin{bmatrix} x_{1t}, y_{t+1}\left|Y_t\right. \\ x_{2t}, y_{t+1}\left|Y_t\right. \\ x_{3t}, y_{t+1}\left|Y_t\right. \end{bmatrix} = P_{t+1|t} \times \mathbb{E}\begin{bmatrix} \overline{\mu}\exp\left(x_{1t}\right)\left|Y_t\right. \\ \overline{\sigma}\exp\left(x_{2t}\right)x_{3t}\left|Y_t\right. \\ \overline{\sigma}\exp\left(x_{2t}\right)\left|Y_t\right. \end{bmatrix},$$

from Stein's Lemma.

6.2.2 Results

As with the stochastic volatility model, it is a big challenge to prove that the parameters of the model are well identified. To get a sense of the precision of Quasi-Maximum-Likelihood (QML) estimates, simple identification exercises were perform. The first study compares the value of the quasi-log likelihood function of the nonlinear state space model under the Taylor series approach for different orders of approximation as well as the exact approach. The parameters used to evaluate the quasi log-likelihood functions were obtained from Brandt and Kang (2004). Figure 4 contains the results. The figure contains in the x axis the different orders of approximation while the y axis contains the values of the quasi-likelihood function.

[Insert Figure 4 about here]

The second exercise consists of simulating a sample path for the stock returns with $T = 5000$ using the parameter estimates obtained by Brandt and Kang (2004) and a degree of approximation of $M = 10$. The likelihood function under each method was evaluated numerically by fixing all the parameter values but one; all of the other parameter values were the true parameter values which were used to simulate the data. The results are shown in Figures 5 ; the dashed lines represent the unknown parameter values used to generate the data. The figures present the quasi log-likelihood function as a function of a_{11}, a_{21}, a_{12} and a_{22}. The concavity of the quasi-likelihood function with respect to each parameter shows evidence that all the parameters are well identified and the maxima are achieved in parameter values close enough to the ones used to simulate the data. For example, the value used to simulate was $a_{11} = 0.8589$;

the QML estimate obtained was 0.9191. It is worth mentioning that the number of simulations is important; with a small number of observations the correlation coefficient ρ, may not be identified.

[Insert Figure 5 about here]

To provide more evidence of the precision of these estimation methods, we simulated 500 independent samples of $T = 792$ monthly returns with parameter values obtained from Brandt and Kang (2004) and estimated repeatedly the parameters with each of the techniques mentioned above. Table 1, Panel A presents the parameter estimates obtained by the QML estimation method. The first column shows the parameter values used to simulate the data and the remaining columns contain the sampling distribution of the corresponding parameter estimates. Overall, the parameter estimates show evidence of consistency; however, the small number of simulations does not allow to check for a more precise assessment of asymptotic results. The standard deviations for b_{11}, b_{22}, $\overline{\mu}$ and $\overline{\sigma}$ are relatively small compared to the overall standard deviations of the other parameters. As for the correlation coefficients ρ and ρ_μ; it is interesting to note that, in general, the parameter estimates of these correlation coefficients are not identified. [20] A common approach to correct for this issue is to add another observation equation or a set of predictors in the dynamics of the mean and volatility of returns.

6.2.3 Data

I study monthly returns on the value weighted CRSP index in excess of the one month Treasury bill rate from January 1946 through December 2011 (792 observations). The short rate is the yield of a one-month Treasury bill. Table 6 presents summary statistics of the data and Figure 3 plots the series.

[Insert Table 6 about here]

6.2.4 Parameter Estimates

Table 7 presents the quasi-maximum likelihood estimates of the latent VAR under the four Model representations. Under , it is assumed that the innovations of the transition and observation equation are uncorrelated. The innovations of the conditional mean and volatility are contemporaneously negative correlated and are statistically significant under both specifications. (The t-statistics were computed with the asymptotic standard errors.) As a result, we

[20] The lack of identification of correlation coefficients is a common problem in the standard filtering applications; see Hamilton (1994) for details.

strongly reject the hypothesis of lack of contemporaneous relationship between the conditional mean and the conditional volatility. In addition, the volatility in mean and mean in volatility are negative and significant. These results are consistent with French, Schwert, and Stambaugh (1987), Campbell and Hentschel (1992) and Brandt and Kang (2004). Under the first estimation method a_{21} is significant and a_{12} is not, in contrast with the approximation, that reflects no significance for a_{12} and a_{21}.

[Insert Table 7 about here]

6.3 A Dynamic Stochastic Equilibrium Model

In this section, I estimate a simple DSGE model. [21] Flury and Shephard (2011) notices that the particle filters are the only feasible approach to estimating parameters of DSGE models; however, the filtering and estimation technique that I propose in this paper is another way to perform estimation.

6.3.1 The Model

There is a representative household maximizing its lifetime utility given by:

$$E_0\left[\sum_{t=0}^{\infty}\beta^t\frac{C_t^{1-\eta}}{1-\eta}\right], \quad \beta\in(0,1), \quad \eta>0 \tag{31}$$

where C_t is consumption at time t, β is the subjective discount factor and η is the risk aversion parameter.

There is one single good produced according to

$$Y_t = A_t K_t^{\alpha}, \tag{32}$$

where K_t is the stock of capital and A_t is the technology. The stock of capital evolves according to

$$K_{t+1} = (1-\delta)\,K_t + I_t,$$

where I_t denotes investment and δ is the depreciation rate. The aggregate resource constraint is

$$C_t = I_t + Y_t.$$

[21] An and Schorfheide (2007) considered Bayesian inference for DSGE models. Fernández-Villaverde and Rubio-Ramírez (2007) used particle filters to perform parameter inference. Flury and Shephard (2011) provide a Bayesian based method based on particle filters.

I assume that

$$\log A_{t+1} = \rho \log A_t + \varepsilon_{t+1}, \tag{33}$$

where $\varepsilon_t \sim N\left(0, \sigma_A^2\right)$. The central planner's problem is

$$\max_{\{C_t, K_{t+1}\}_{t=0}^{\infty}} E_0 \left[\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\eta}}{1-\eta} \right], \qquad \beta \in (0,1), \qquad \eta > 0 \tag{34}$$

subject to

$$K_{t+1} + C_t \leq A_t K_t^{\alpha} + (1-\delta) K_t$$

and (32), for $t = 0, 1, ...;$ K_0, A_0 given.

The first order conditions implied by (33), (34) and (32) are

$$\begin{aligned} C_t^{-\eta} &= \beta E_t \left[C_{t+1}^{-\eta} \left(1 - \delta + \alpha A_{t+1} K_{t+1}^{\alpha-1} \right) \right], \\ K_{t+1} &= A_t K_t^{\alpha} + (1-\delta) K_t - C_t, \\ \log A_{t+1} &= \rho \log A_t + \varepsilon_{t+1}. \end{aligned} \tag{35}$$

These equations fully characterize the solution to the optimization problem faced by the central planner.

The solution to the system in (35) involves finding policy functions g and h such that

$$\begin{aligned} C_t &= g\left(K_t, A_t\right) \\ \begin{bmatrix} K_{t+1} \\ \log A_{t+1} \end{bmatrix} &= \begin{bmatrix} h_1\left(K_t, A_t\right) \\ h_2\left(K_t, A_t\right) \end{bmatrix} + \sigma \begin{bmatrix} 0 \\ \sigma_A \end{bmatrix} \varepsilon_{t+1} \end{aligned}$$

where σ is a perturbation parameter. This system of equilibrium equations does not have a general analytical solution, and I solved it with a numerical method. The second-order approximation to these policy functions is obtained via perturbation methods.[22] The system is solved in terms of log-deviations from a non-stochastic steady state. Let

$$\begin{aligned} \widehat{c}_t &= \log\left(C_t / C_{ss}\right) \\ \widehat{k}_t &= \log\left(K_t / K_{ss}\right) \\ \widehat{a}_t &= \log\left(A_t\right) \end{aligned}$$

where C_{ss} and K_{ss} are the non-stochastic steady state values for C_t and K_t, respectively:

[22] See Judd (1998) for a detailed explanation of perturbation methods in Economics.

$$
\begin{aligned}
C_{ss} &= (K_{ss})^{\alpha} - \delta K_{ss}, \\
K_{ss} &= \left[\frac{\alpha\beta}{1-\beta(1-\delta)}\right]^{\frac{1}{1-\alpha}}.
\end{aligned}
$$

For simplicity, let $\widehat{x}_t = \left[\widehat{k}_t, \widehat{a}_t\right]'$, denote the log-deviations from the non-stochastic steady state. The solution of the system will be of the state form

$$
\begin{aligned}
\widehat{k}_{t+1} &= h_{x,1}\widehat{x}_t + \frac{1}{2}\widehat{x}_t' h_{xx,1}\widehat{x}_t + \frac{1}{2}h_{\sigma\sigma,1}\sigma^2, \\
\widehat{a}_{t+1} &= \rho\widehat{a}_t + \varepsilon_{t+1},
\end{aligned} \tag{36}
$$

and

$$
\widehat{c}_t = g_x\widehat{x}_t + \frac{1}{2}\widehat{x}_t' g_{xx,1}\widehat{x}_t + \frac{1}{2}g_{\sigma\sigma}\sigma^2
$$

I rely on the code from Schmitt-Grohe and Uribe (2004) to solve for the unknown derivatives $h_x, h_{xx}, h_{\sigma\sigma}, g_x, g_{xx}$ and $g_{\sigma\sigma}$.

6.3.2 State Space representation and Implementation

Following Flury and Shephard (2011), I assume that the observable variable is the detrended real gross domestic product per capita, $\widehat{GDP}_t$

$$
\widehat{GDP}_t = \widehat{y}_t + \sigma_y\varepsilon_t^y. \tag{37}
$$

From Equation (32), the log-GDP is given by

$$
\widehat{y}_t = \widehat{a}_t + \alpha\widehat{k}_t.
$$

Finally, Eqs. (36) and (37) specify a non-linear state space system

$$
\begin{aligned}
\widehat{GDP}_t &= \widehat{a}_t + \alpha\widehat{k}_t + \sigma_y\varepsilon_t^y, \\
\widehat{a}_{t+1} &= \rho\widehat{a}_t + \varepsilon_{t+1}, \\
\widehat{k}_{t+1} &= h_{x,1}\widehat{x}_t + \frac{1}{2}\widehat{x}_t' h_{xx,1}\widehat{x}_t + \frac{1}{2}h_{\sigma\sigma,1}\sigma^2
\end{aligned} \tag{38}
$$

where the last equation can be rewritten as

$$
\widehat{k}_{t+1} = h_k\widehat{k}_t + h_a\widehat{a}_t + \frac{1}{2}\left(h_{kk}\widehat{k}_t^2 + 2h_{ak}\widehat{a}_t\widehat{k}_t + h_{aa}\widehat{a}_t^2\right) + \frac{1}{2}h_{\sigma\sigma,1}\sigma^2. \tag{39}
$$

6.3.3 Quasi-Likelihood Function

In order to approximate the likelihood of the model, $L(\theta)$, the first and second moments of equations (37) and (38) are necessary. In this case, the transition equation in (38) is quadratic in the state variables; for that reason, the second order Taylor series expansion of the transition equation will coincide with the function value for all values of x, and the Taylor series approach to evaluate the mean and variance of the transition equation is exact. The mean vector of state variables,

$$x_{t+1|t} = \left[\ \mathbb{E}\left[\widehat{k}_{t+1} \left| Y_t \right.\right] \quad \mathbb{E}\left[\widehat{a}_{t+1} \left| Y_t \right.\right] \ \right]'$$

is computed by applying the second order Taylor series approximations of (39) from Section 4. Now, since the transition equation is quadratic, its variance requires fourth order polynomials, as a result, the variance of the transition is computed with a fourth order Taylor series, as

$$P_{t+1|t} = \begin{bmatrix} Var_t\left[\widehat{k}_{t+1} \left| Y_t \right.\right] & cov_t\left[\widehat{k}_{t+1}, \widehat{a}_{t+1} \left| Y_t \right.\right] \\ cov_t\left[\widehat{k}_{t+1}, \widehat{a}_{t+1} \left| Y_t \right.\right] & Var_t\left[\widehat{a}_{t+1} \left| Y_t \right.\right] + \sigma_A^2 \end{bmatrix}.$$

The observation equation is linear in the state variables, as a result, its first and second conditional moments are

$$y_{t+1|t} = \mathbb{E}\left[y_{t+1} \left| Y_t \right.\right] = [1, \alpha] \cdot x_{t+1|t}$$

and

$$P_{t+1|t}^{yy} = [1, \alpha] \cdot P_{t+1|t} \cdot [1, \alpha]' + \sigma_y^2.$$

Finally, the covariance between the observation and transition equation is

$$P_{t+1|t}^{xy} = P_{t+1|t} \cdot [1, \alpha]'.$$

The task for the econometrician is to carry out inference on $\theta = (\alpha, \beta, \delta, \eta, \rho, \sigma_A, \sigma_y)$. The algorithm to obtain $L(\theta)$ works by computing C_{ss} and K_{ss} given the first choice of parameter values, $\theta^{(i)}$, and then use perturbation methods to find numerical values for $h_x, h_{xx}, h_{\sigma\sigma}, g_x, g_{xx}$ and $g_{\sigma\sigma}$, and then run the Taylor-Kalman filter to obtain $L\left(\theta^{(i)}\right)$.

6.3.4 Results

The task for the econometrician is to carry out inference on $\theta = (\alpha, \beta, \delta, \eta, \rho, \sigma_z, \sigma_y)$, where σ_y is the standard deviation of the measurement error. I use the model to simulate an economy with $T = 500$ and the following parameterization:

β	δ	α	ρ	γ	σ_y	σ_ε
0.95	0.15	0.30	0.9	3	0.3	0.2

The quality of the filtering algorithm based on Taylor approximations was tested using Monte Carlo simulations. A sample path of size T was simulated using the parameter values described in the previous table. Using an approximation of $M = 4$, the filtering recursions were calculated and the state estimates were estimated. [23]

The results for the filtered state variables are shown in Figure 7. In this case, I compare the simulated Log-investments and shocks with the filtered shocks and filtered Log-investment. The difference between the simulated paths and the filtered ones is almost indistinguishable.

[Insert Figure 7 about here]

As for the parameter estimation part, a basic identification exercise was performed. A sample path of random shocks was simulated with $T = 500$ using the previous parameterization. As in the simulation, the degree of approximation of the Quasi-likelihood function is $M = 4$. For each simulation, the likelihood function was evaluated numerically by fixing all the parameter values but the one showed in the $x-$axis. The results are shown in Figures 8 and 9.

[Insert Figures 8 and 9 about here]

As reported previously by Flury and Shephard (2011) some of the parameter values are not well identified, such as the subjective discount factor, β, the depreciation rate,δ, and there is some bias in the risk aversion parameter. In this case, the risk aversion parameter was identified, but the value estimated via QML is close to 0.5. All the other parameters, such as the volatility of shocks, measurements, correlations and α are identified. An alternative way to achieve identification by included other sets of observables such as investment or noisy measures of consumption growth. This exercises are left for future work.

7 Conclusions

In this paper, I proposed a new nonlinear filter based on Taylor series approximations. This filter is able to perform state and parameter estimation when the process and measurement models are highly nonlinear. My estimation results suggest that filtering methods via the higher order Taylor Series filter are superior to conventional methods such as the Extended Kalman Filter or the Unscented Kalman Filter. In terms of computational efficiency, my filter outperforms the standard particle filter techniques. The filter can be applied directly application

[23] The choice of $M = 4$ is due to the second order approximation used to solve the policy function and the fact that these filtering recursion involved first and second moments.

in stochastic volatility models, predictive systems and structural estimation among others. My results suggest that this filter may be a good a approach for a number of problems that involve nonlinear dynamic modeling in Finance and Economics.

References

An, S., F. Schorfheide, 2007. Bayesian Analysis of DSGE Models. Econometric Reviews 26(2-4), 113–172.

Andersen, T. G., T. Bollerslev, F. X. Diebold, H. Ebens, 2001. The distribution of realized stock return volatility. Journal of Financial Economics 61(1), 43 – 76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, P. Labys, 2003. Modeling and Forecasting Realized Volatility.. Econometrica 71(2), 579.

Andersen, T. G., B. E. Sørensen, 1996. GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study. Journal of Business and Economic Statistics 14(3), pp. 328–352.

Bansal, R., A. Yaron, 2004. Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. The Journal of Finance 59(4), 1481–1509.

Barillas, F., L. P. Hansen, T. J. Sargent, 2009. Doubts or variability?. Journal of Economic Theory 144(6), 2388 – 2418.

Bidder, R., M. Smith, 2011. Doubts and Variability. New York University Working Paper.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, S. Terry, 2012. Really Uncertain Business Cycles. Working Paper: Stanford University.

Boguth, O., M. Carlson, A. Fisher, M. Simutin, 2011. Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas. Journal of Financial Economics 102(2), 363 – 389.

Bollerslev, T., J. M. Wooldridge, 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11(2), 143–172.

Brandt, M. W., Q. Kang, 2004. On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach. Journal of Financial Economics 72(2), 217 – 257.

Broto, C., E. Ruiz, 2004. Estimation Methods for Stochastic Volatility Models: a Survey.. Journal of Economic Surveys 18(5), 613 – 649.

Calvet, L., A. J. Fisher, L. Wu, 2010. Dimension-Invariant Dynamic Term Structures. Working Paper: University of British Columbia.

Campbell, J., A. Sunderam, L. Viceira, 2011. Inflation bets or deflation hedges? the changing risks of nominal bonds. Working Paper: Harvard University.

Campbell, J. Y., L. Hentschel, 1992. No news is good news: An asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31(3), 281 – 318.

Christoffersen, P., K. Jacobs, L. Karoui, K. Mimouni, 2012. Nonlinear Kalman Filtering in Affine Term Structure Models. Working Paper: University of Toronto.

Clark, T., 2009. Is the Great Moderation over? An empirical analysis. Economic Review 4, 5–42.

Crisan, D., B. Rozovskii, 2011. The Oxford handbook of Nonlinear Filtering. Oxford University Press, .

Croushore, D., T. Stark, 2001. A real-time data set for macroeconomists. Journal of Econometrics 105(1), 111 – 130.

Doucet, A., N. De Freitas, N. Gordon, 2001. Sequential Monte Carlo methods in practice. Springer Verlag, .

Fama, E. F., K. R. French, 1988. Permanent and Temporary Components of Stock Prices. Journal of Political Economy 96(2), pp. 246–273.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, 2007. Estimating Macroeconomic Models: A Likelihood Approach.. Review of Economic Studies 74(4), 1059 – 1087.

Flury, T., N. Shephard, 2011. Bayesian Inference based only on Simulated Likelihood: Particle Filter Analysis of Dynamic Economic Models. Econometric Theory 27(Special Issue 05), 933–956.

French, K. R., G. Schwert, R. F. Stambaugh, 1987. Expected stock returns and volatility. Journal of Financial Economics 19(1), 3 – 29.

Gallant, A., H. White, 1988. A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Basil Blackwell New York, .

Garlappi, L., G. Skoulakis, 2010. Solving Consumption and Portfolio Choice Problems: The State Variable Decomposition Method. Review of Financial Studies 23(9), 3346–3400.

——— , 2011. Taylor series approximations to expected utility and optimal portfolio choice. Mathematics and Financial Economics 5, 121–156.

Ghysels, E., A. C. Harvey, E. Renault, 1996. Stochastic volatility vol. 14. Elsevier North-Holland, .

Hamilton, J. D., 1994. Time Series Analysis. Princeton University Press, .

Ito, K., K. Xiong, 2000. Gaussian Filters for Non-Linear Filtering Problems. Automatic Control, IEEE Transactions on 45(5), 910 –927.

Jacquier, E., M. Johannes, N. Polson, 2007. MCMC maximum likelihood for latent state models. Journal of Econometrics 137(2), 615 – 640.

Jacquier, E., N. G. Polson, P. E. Rossi, 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. Journal of Econometrics 122(1), 185 – 212.

Jazwinski, A. H., 1970. Stochastic Processes and Filtering Theory. Academic Press, .

Judd, K. L., 1998. Numerical Methods in Economics. MIT Press (Cambridge, Mass.), .

Julier, S. J., J. K. Uhlmann, 1997. A New Extension of the Kalman Filter to Nonlinear Systems. Proceedings of AeroSense: 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls, Orlando, FL.

——— , 2004. Unscented Filtering and Nonlinear Estimation. Proceedings of the IEEE 92(3), 401 – 422.

Justiniano, A., G. E. Primiceri, 2008. The Time-Varying Volatility of Macroeconomic Fluctuations. American Economic Review 98(3), 604–41.

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME–Journal of Basic Engineering 82(Series D), 35–45.

Kim, C.-J., C. R. Nelson, 1999. State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications vol. 1. the MIT Press, .

Kim, S., N. Shephard, S. Chib, 1998. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. The Review of Economic Studies 65(3), 361–393.

Lamoureux, C., G. Zhou, 1996. Temporary components of stock returns: What do the data tell us?. Review of Financial Studies 9(4), 1033–1059.

Ludvigson, S., 2012. Advances in Consumption-Based Asset Pricing. Working Paper: New York University.

McConnell, M. M., G. Perez-Quiros, 2000. Output Fluctuations in the United States: What Has Changed Since the Early 1980's?. The American Economic Review 90(5).

Newey, W., K. West, 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55(3), 703–708.

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

Savits, T. H., 2006. Some statistical applications of Faa di Bruno. Journal of Multivariate Analysis 97(10), 2131 – 2140.

Schmitt-Grohe, S., M. Uribe, 2004. Solving dynamic general equilibrium models using a second-order approximation to the policy function. Journal of Economic Dynamics and Control 28(4), 755–775.

Shephard, N. E., 2005. Stochastic Volatility: Selected Readings. Oxford University Press, USA, .

Stein, C. M., 1981. Estimation of the Mean of a Multivariate Normal Distribution. The Annals of Statistics 9(6), 1135–1151.

Stock, J., M. Watson, 2002. Has the Business Cycle Changed and Why?. NBER Working Papers.

Tallarini, T. D. J., 2000. Risk-sensitive real business cycles. Journal of Monetary Economics 45(3), 507 – 532.

Tanizaki, H., R. Mariano, 1996. Nonlinear filters based on taylor series expansions?. Communications in Statistics-Theory and Methods 25(6), 1261–1282.

Taylor, S., 2008. Modelling Financial Time Series. World Scientific, .

van Binsbergen, J., R. Koijen, 2011. Likelihood-based Estimation of Exactly-Solved Present-Value Models. Working Paper: University of Chicago.

White, H., 1982. Maximum Likelihood Estimation of Misspecified Models. Econometrica 50(1), pp. 1–25.

Winschel, V., M. Krätzig, 2010. Solving, Estimating, and Selecting Nonlinear Dynamic Models Without the Curse of Dimensionality. Econometrica 78(2), 803–821.

A Efficient Calculation of derivatives of composite functions

The efficient computation of partial derivatives relies on the Taylor expansion of a function of the form $f(\underline{x}) = h(g(\underline{x}))$, where $h : \mathbb{R} \longrightarrow \mathbb{R}$, $g : \mathbb{R}^N \longrightarrow \mathbb{R}$, and $\underline{x} = (x_1, x_2, ..., x_N)$ denotes an $N-$dimensional vector[24]. The generic $M - th$ order Taylor expansion of f centered at a constant point $\mathbf{0}_N$ is

$$f(\underline{x}) \simeq \sum_{\{\mathbf{q}:|\mathbf{q}|\leq M\}} \frac{1}{\mathbf{q}!} f_{\mathbf{q}}(\mathbf{0}_N) \prod_{n=1}^{N} x_n^{q_n} \tag{40}$$

where $\mathbf{q} = (q_1, ..., q_N)$ is a vector of nonnegative integers, $|\mathbf{q}| = \sum_{n=1}^{N} q_n$, $\mathbf{q}! = \prod_{n=1}^{N}(q_n!)$, and $f_{\mathbf{q}}(\mathbf{0}_N)$ denotes the partial derivative of order $\mathbf{q}$ of the function $f(\underline{x})$ evaluated at $\mathbf{0}_N$, i.e.,

$$f_{\mathbf{q}}(\mathbf{0}_N) = \frac{\partial^{q_1+...+q_N} f}{\partial x_1^{q_1} ... \partial x_1^{q_N}}(\mathbf{0}_N). \tag{41}$$

To compute such derivatives, Savits (2006) relies on the recursive formula of Faà di Bruno (1855, 1857). To present the formula, I will introduce some notation. Let $\mathbb{N}_0$ denote the set of nonnegative integers and let $\mathbf{q} = (q_1, ..., q_N)$, where $q_1, ..., q_N$ are nonnegative integers. We write $\ell \leq \mathbf{q}$ if $\ell_n \leq q_n$, for $n = 1, ..., N$, and denote

$$\binom{\mathbf{q}}{\ell} = \frac{\mathbf{q}!}{\ell!(\mathbf{q}-\ell)!}.$$

Let $g_{\mathbf{q}}(\underline{x})$ denote the partial derivative of order $\mathbf{q}$ of the function $g(\underline{x})$, and $h_n(w)$ denote the $n-$th derivative of the function $h(w)$ with respect to the one dimensional variable w. According to the multivariate version of Faà di Bruno's formula, the partial derivative of order $\mathbf{q}$ of the composite function $f(x) = h(g(\underline{x}))$, i.e., $f_{\mathbf{q}}(\underline{x})$, can be expressed as

$$f_{\mathbf{q}}(\underline{x}) = \sum_{n=1}^{|\mathbf{q}|} h_n(g(\underline{x})) \alpha_{\mathbf{q},n}(\underline{x}), \tag{42}$$

where $\alpha_{\mathbf{q},n}(\underline{x})$ are homogeneous polynomials of degree n in the partial derivatives of g, $g_\ell(\underline{x})$, $\ell \leq \mathbf{q}$. To compute the generic derivative of f, it is, therefore, sufficient to determine the polynomials $\alpha_{\mathbf{q},n}(\underline{x})$. These can be computed efficiently by relying on the recursive relationship proved in Theorem 3.1 of Savits (2006).

[24] For simplicity I consider the the case in which $f(x)$ is one-dimensional; however, the formulas can be extended directly to a multi-dimensional case by applying the results for the one-dimensional case to each of the components.

Theorem A.1 *For $\mathbf{q} \geq \mathbf{0}_N$, $1 \leq j \leq N$, and $1 \leq n \leq |\mathbf{q}| + 1$, we have*

$$\alpha_{q+e_j,n}\left(\underline{x}\right) = \sum_{\left\{\ell \in \mathbb{N}_0^N:\ \mathbf{0}_N \leq \ell \leq \mathbf{q}; |\ell| \geq n-1\right\}} \binom{\mathbf{q}}{\mathbf{l}} g_{\mathbf{q}+\mathbf{e}_j-\ell}\left(x\right) \alpha_{\ell,n-1}\left(\underline{x}\right) \tag{43}$$

where $\mathbf{e}_j$ is the unit vector with $j-th$ component equal to 1 and we set

$$\alpha_{\ell,\mathbf{0}}\left(\underline{x}\right) = \begin{cases} 1 \text{ if } \ell = \mathbf{0}, \\ 0 \text{ if } \ell \neq \mathbf{0}, \end{cases}$$

Proof *See Savits (2006).*

Proposition A.2 *Faa di Bruno Formula. Let $f\left(\mathbf{x}\right) = g\left(\mathbf{x}\right)^2$, then by the Faa di Bruno formula for the derivatives of a compound function, we have that the derivatives of $h\left(x\right)$, are given by*

$$\begin{aligned} f_{\mathbf{0}}\left(\mathbf{x}\right) &= g\left(\mathbf{x}\right)^2 \\ f_{\mathbf{q}+\mathbf{e}_j}\left(\mathbf{x}\right) &= 2 \sum_{\left\{\ell \in \mathbb{N}_0^N:\ \mathbf{0}_N \leq \ell \leq \mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} g_{\mathbf{q}+\mathbf{e}_j-\ell}\left(\mathbf{x}\right) g_{\ell}\left(\mathbf{x}\right) \end{aligned} \tag{44}$$

Proof For the Faa di Bruno, we take $h\left(y\right) = y^2$, then $h_0\left(y\right) = y^2$, $h_1\left(y\right) = 2y$ and $h_2\left(y\right) = 2$. Then, by the Faa di Bruno we have

$$\begin{aligned} f_{\mathbf{q}}\left(\mathbf{x}\right) &= h_1\left(g\left(\mathbf{x}\right)\right) \alpha_{\mathbf{q},1}\left(\mathbf{x}\right) + h_2\left(g\left(\mathbf{x}\right)\right) \alpha_{\mathbf{q},2}\left(\mathbf{x}\right) \\ &= 2g\left(\mathbf{x}\right) \alpha_{\mathbf{q},1}\left(\mathbf{x}\right) + 2\alpha_{\mathbf{q},2}\left(\mathbf{x}\right). \end{aligned}$$

Now, applying (43) to $n = 1, 2$, we have

$$\begin{aligned} \alpha_{\mathbf{q}+\mathbf{e}_j,1}\left(\underline{x}\right) &= \sum_{\left\{\ell \in \mathbb{N}_0^N:\ \mathbf{0}_N \leq \ell \leq \mathbf{q}; |\ell| \geq 0\right\}} \binom{\mathbf{q}}{\ell} g_{\mathbf{q}+\mathbf{e}_j-\ell}\left(x\right) \alpha_{\ell,0}\left(\underline{x}\right) \\ &= g_{\mathbf{q}+\mathbf{e}_j}\left(x\right) \end{aligned}$$

$$\alpha_{\mathbf{q}+\mathbf{e}_j,2}\left(\underline{x}\right) = \sum_{\left\{\ell \in \mathbb{N}_0^N:\ \mathbf{0}_N \leq \ell \leq \mathbf{q}; |\ell| \geq 1\right\}} \binom{\mathbf{q}}{\ell} g_{\mathbf{q}+\mathbf{e}_j-\ell}\left(x\right) \alpha_{\ell,1}\left(\underline{x}\right).$$

Finally,

$$
\begin{aligned}
f_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x}) &= 2g(\mathbf{x})\,\alpha_{\mathbf{q}+\mathbf{e_j},1}(\mathbf{x}) + 2\alpha_{\mathbf{q}+\mathbf{e_j},2}(\mathbf{x}) \\
&= 2g(\mathbf{x})\,g_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x}) + 2\sum_{\left\{\ell\in\mathbb{N}_0^N:\ \mathbf{0}_N\leq\ell\leq\mathbf{q};|\ell|\geq 1\right\}} \binom{\mathbf{q}}{\ell} g_{\mathbf{q}+\mathbf{e}_j-\ell}(x)\,g_{\ell,1}(\underline{x}) \\
&= \sum_{\left\{\ell\in\mathbb{N}_0^N:\ \mathbf{0}_N\leq\ell\leq\mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} 2\cdot g_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\,g_{\ell}(\mathbf{x})
\end{aligned}
$$

Corollary A.3 *Let $f(\mathbf{x}) = g_1(\mathbf{x})\,g_2(\mathbf{x})$, $g_1, g_2 : \mathbb{R}^N \longrightarrow \mathbb{R}$ then the derivatives of $h(x)$, are given by*

$$
\begin{aligned}
f_{\mathbf{0}}(\mathbf{x}) &= g^1(\mathbf{x})\,g^2(\mathbf{x}) \qquad (45)\\
f_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x}) &= \sum_{\left\{\ell\in\mathbb{N}_0^N:\ \mathbf{0}_N\leq\ell\leq\mathbf{q}\right\}} \binom{\mathbf{q}}{\mathbf{l}} g^1_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\,g^2_{\ell}(\mathbf{x}) \\
&\quad + \sum_{\left\{\ell\in\mathbb{N}_0^N:\ \mathbf{0}_N\leq\ell\leq\mathbf{q}\right\}} \binom{\mathbf{q}}{\mathbf{l}} g^2_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\,g^1_{\ell}(\mathbf{x})
\end{aligned}
$$

Proof *This proof is a direct consequence from the following algebraic identity*

$$
\left(g^1 g^2\right)(\mathbf{x}) = \frac{\left(g^1+g^2\right)^2 - \left(g^1\right)^2 - \left(g^2\right)^2}{2}(\mathbf{x}).
$$

Hence,

$$
\left(g^1 g^2\right)_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x}) = \left(\frac{\left(g^1+g^2\right)^2 - \left(g^1\right)^2 - \left(g^2\right)^2}{2}\right)_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x})
$$

From the previous lemma, we know that the derivatives for $\left(g^1(x)+g^2(x)\right)^2$ *satisfy:*

$$\begin{aligned}
\left(g^1+g^2\right)^2_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x}) &= \sum_{\left\{\ell\in\mathbb{N}_0^N:\ \mathbf{0}_N\leq\ell\leq\mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} 2\cdot\left(g^1+g^2\right)_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\left(g^1+g^2\right)_{\ell}(\mathbf{x}) \\
&= \sum_{\left\{\ell\in\mathbb{N}_0^N:\ \mathbf{0}_N\leq\ell\leq\mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} 2\cdot g^1_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\, g^1_{\ell}(\mathbf{x}) \\
&\quad + \sum_{\left\{\ell\in\mathbb{N}_0^N:\ \mathbf{0}_N\leq\ell\leq\mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} 2\cdot g^2_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\, g^2_{\ell}(\mathbf{x}) \\
&\quad + \sum_{\left\{\ell\in\mathbb{N}_0^N:\ \mathbf{0}_N\leq\ell\leq\mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} 2\cdot g^1_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\, g^2_{\ell}(\mathbf{x}) \\
&\quad + \sum_{\left\{\ell\in\mathbb{N}_0^N:\ \mathbf{0}_N\leq\ell\leq\mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} 2\cdot g^2_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\, g^1_{\ell}(\mathbf{x}) \\
&= \left(g^1\right)^2_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x}) + \left(g^2\right)^2_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x}) \\
&\quad + \sum_{\left\{\ell\in\mathbb{N}_0^N:\ \mathbf{0}_N\leq\ell\leq\mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} 2\cdot g^1_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\, g^2_{\ell}(\mathbf{x}) \\
&\quad + \sum_{\left\{\ell\in\mathbb{N}_0^N:\ \mathbf{0}_N\leq\ell\leq\mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} 2\cdot g^2_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\, g^1_{\ell}(\mathbf{x})
\end{aligned}$$

Now, subtracting $(g^1)^2_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x})$, $(g^2)^2_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x})$ *and dividing by two, yields to the desired result.*

B Stein's Lemmas

Lemma B.1 *Let $Z \sim N\left(\mu, \sigma^2\right)$ and let f any continuously differentiable function such that f' exists almost everywhere and $E\left|f'(Z)\right| < \infty$, then*

$$cov\left(Z, f(Z)\right) = E\left[(Z-\mu) f(Z)\right] = \sigma^2 E\left[f'(Z)\right].$$

Proof

$$\begin{aligned} E\left[(Z-\mu) f(Z)\right] &= \int_{-\infty}^{\infty} (z-\mu) f(z) \frac{e^{-\frac{(z-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} dz \\ &= -\sigma^2 f(z) \frac{e^{-\frac{(z-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \Bigg|_{-\infty}^{\infty} + \sigma^2 \int_{-\infty}^{\infty} f'(z) \frac{e^{-\frac{(z-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} dz \\ &= \sigma^2 E\left[f'(Z)\right]. \end{aligned}$$

Proposition B.2 *For any function $f(x_1, ..., x_N)$ such that $\partial f / \partial x_i$ exists almost everywhere and $E\left|\frac{\partial}{\partial x_i} f(X)\right| < \infty$, $i = 1, ..., n$. Let $\nabla f(X) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)^{\top}$. Then the following identity holds*

$$cov\left(X, f(X)\right) = \Sigma \times E\left[\nabla f(X)\right], \tag{46}$$

more specifically

$$cov\left(X_1, f(X_1, ..., X_N)\right) = \sum_{i=1}^{N} cov\left(X_1, X_i\right) \times E\left[\frac{\partial}{\partial x_i} f(X_1, ..., X_N)\right].$$

Proof Let $\mathbf{Z} = (Z_1, ..., Z_N)$, where Z_i are i.i.d. $N(0,1)$ random variables. From the previous lemma, we know that for any $g(X)$, differentiable almost everywhere, $cov\left[Z_i, g(\mathbf{Z})\right] = E\left[\partial g / \partial z_i\right]$. Hence

$$cov\left[\mathbf{Z}, g(\mathbf{Z})\right] = E\left[\nabla g(\mathbf{Z})\right]. \tag{47}$$

Stein (1981) provides a more elaborated proof of the previous result. Now, the random vector X can be written as $X = \Sigma^{1/2}\mathbf{Z}+\mu$, and $f(\mathbf{Z}) = g\left(\Sigma^{1/2}\mathbf{Z}+\mu\right)$. Hence, the left hand side of (46) is

$$\begin{aligned} cov\left[\mathbf{X}, f(\mathbf{X})\right] &= cov\left[\Sigma^{1/2}\mathbf{Z}+\mu, g\left(\Sigma^{1/2}\mathbf{Z}+\mu\right)\right] \\ &= \Sigma^{1/2} cov\left[\mathbf{Z}, g\left(\Sigma^{1/2}\mathbf{Z}+\mu\right)\right] \\ &= \Sigma E\left[\nabla f(\mathbf{X})\right]. \end{aligned}$$

C Standard Kalman Filter

The state space representation of a linear model is given by

$$\begin{aligned} y_t &= Hx_t + v_t \\ x_t &= Fx_{t-1} + w_t \end{aligned} \tag{48}$$

where $v_t \sim N(0, R)$ and $w_t \sim N(0, Q)$. The first equation is the observation equation and represents the true measurement of the state variable x_t, and H represents the model that maps the true state space into the observed space; v_t is the measurement noise. The second equation represents the evolution of the state variable of the state variable.

In systems like (48) where the state variables are normally distributed and the measurement equations are linear, the standard Kalman filter yields to efficient state estimates in a minimum variance criteria. The estimates can be obtained using the Kalman filter update and prediction rules. Following, Kalman (1960), the optimal estimate of $\widehat{x}_{t+1|t+1.}$ (in an minimum variance sense) is given by updating the prediction equation with the current measurement.

A prediction state is given by

$$\begin{aligned} \widehat{x}_{t+1|t} &= F\widehat{x}_{t|t} \\ P_{t+1|t} &= FP_{t|t}F^T + Q. \end{aligned} \tag{49}$$

The update rule is given by

$$\begin{aligned} \widehat{x}_{t+1|t+1} &= \widehat{x}_{t+1|t} + K_{t+1}\left(y_{t+1} - H\widehat{x}_{t+1|t}\right) \\ P^{yy}_{t+1|t} &= HP_{t+1|t}H^T + R \\ K_{t+1} &= P_{t+1|t}H^T\left[P^{yy}_{t+1|t}\right]^{-1} \\ P_{t+1|t+1} &= \left(I - K_{t+1}H\right)P_{t+1|t} \end{aligned} \tag{50}$$

D The Extended Kalman Filter

A well known approximation to non linear filtering is the extended Kalman filter, which relies on a first order Taylor expansion of the measurement and transition equations around the predicted value of the state variable at time $x_{t+1|t}$. The measurement equation is written as follows

$$y_{t+1} = h\left(x_{t+1|t}\right) + H_{t+1}\left(x_{t+1} - x_{t+1|t}\right) + v_{t+1} \tag{51}$$

where

$$H_{t+1} = \left.\frac{\partial h}{\partial x_{t+1}}\right|_{x_{t+1}=x_{t+1|t}} \tag{52}$$

denotes the Jacobian matrix of the nonlinear function g computed at $x_{t+1|t}$. The transition equation is linearized as in (52) and is written as

$$x_{t+1} = g\left(x_{t|t}\right) + G_t\left(x_t - x_{t|t}\right) + \varepsilon_{t+1}, \tag{53}$$

where

$$G_t = \left.\frac{\partial g}{\partial x_t}\right|_{x_t=x_{t|t}}.$$

The covariance matrices $P^{xy}_{t+1|t}$ and $P^{yy}_{t+1|t}$ are then computed as

$$P^{xy}_{t+1|t} = P^{xx}_{t+1|t} H_{t+1}, \tag{54}$$

$$P^{yy}_{t+1|t} = H_{t+1} P^{xx}_{t+1|t} H^{\top}_{t+1} + R \tag{55}$$

and

$$P^{xx}_{t+1|t} = G_t P^{xx}_{t|t} G_t + Q$$

The estimate of the state vector is then updated using the standard Kalman filter recursions.

E The Unscented Kalman Filter

The unscented Kalman filter, UKF hereafter, uses the exact nonlinear functions in the observation and transition equations to approximate the moments of the state variables. Unlike the extended Kalman filter, the UKF does not rely on linearizations. The UKF approximates the conditional distribution of the state variables using the unscented transformation Julier and Uhlmann (1997), which is a method for computing statistics of nonlinear transformations of random variables. Julier and Uhlmann (2004) prove that this approximation is accurate to the third order for Gaussian random variables and up to a second order for non-Gaussian states. Moreover, the UKF does not rely on the calculation of Jacobians or Hessian matrices and its efficiency is comparable to the extended Kalman filter as noted by van Binsbergen and Koijen (2011) and Christoffersen, Jacobs, Karoui, and Mimouni (2012).

Let x denote a random vector with mean μ_x and covariance matrix P^{xx}. Consider a nonlinear transformation $y = h(x)$. The basic idea behind the scaled transformation is to generate a set of points, denoted as sigma points, with first and second moments denoted by μ_x and P^{xx}, respectively, and apply the nonlinear transformation to each sigma point. More precisely, the $n-$dimensional random vector is approximated by a set of $2n+1$ weighted points given by

$$\mathcal{X}_0 = \mu_x, \tag{56}$$

$$\mathcal{X}_i = \mu_x + \left(\sqrt{(n+\xi)\,P^{xx}}\right)_i, \text{ for } i = 1, \ldots, n \tag{57}$$

$$\mathcal{X}_i = \mu_x - \left(\sqrt{(n+\xi)\,P^{xx}}\right)_i, \text{ for } i = n+1, \ldots, 2n \tag{58}$$

with weights

$$\begin{aligned} W_0^m &= \frac{\xi}{(n+\xi)}, \\ W_0^c &= \frac{\xi}{(n+\xi)} + \left(1-\rho^2+\theta\right) \\ W_i^m &= W_i^c = \frac{1}{2\,(n+\xi)}, \text{ for } i = 1, ..., n, \end{aligned}$$

where $\xi = \rho^2(n+\lambda) - n$, and where $\left(\sqrt{(n+\xi)\,P^{xx}}\right)_i$ is the $i-$th column of the matrix square root of $(n+\xi)\,P^{xx}$, ρ is a positive scaling parameter that minimizes higher order effects and can be chosen to be arbitrarily small, λ is a positive parameter that guarantees positive-definiteness of the covariance matrix, θ is a nonnegative parameter that can be used to capture higher order moments of the distribution of the state variable. Julier and Uhlmann (1997) propose to use $\theta = 2$ for Gaussian distributions. Once the sigma points are computed, the non linear transformation is applied to each of the sigma points defined in (56) – (58)

$$\mathcal{Y}_i = h\left(\mathcal{X}_i\right), \text{ for } i = 0, ..., n.$$

The unscented Kalman filter relies on the unscented transformation to approximate the covariance matrices $P_{t+1|t}$, $P^{xy}_{t+1|t}$, $P^{yy}_{t+1|t}$. An augmented state vector defined by including the state and measurement noises yielding to a $N_a = 2p + n$ dimensional vector

$$\mathcal{X}^a_t = \begin{bmatrix} x_t \\ \varepsilon_t \\ v_t \end{bmatrix},$$

and the unscented transformation is applied to $\mathcal{X}^a_t$. The process for computing the UKF is summarized as follows:

1. Compute $2N_a + 1$ sigma points of the augmented state space:

$$\begin{aligned} \mathcal{X}^a_{t|t} &= x_{t|t}, \\ \mathcal{X}^a_{t|t} &= x_{t|t} + \left(\sqrt{\left(N_a + \xi\right) P^a_{t|t}}\right)_i, \text{ for } i = 1, \ldots, N_a \\ \mathcal{X}^a_{t|t} &= x_{t|t} - \left(\sqrt{\left(N_a + \xi\right) P^a_{t|t}}\right)_i, \text{ for } i = N_a + 1, \ldots, 2N_a \end{aligned} \tag{59}$$

2. Prediction Step:

$$\begin{aligned} \mathcal{X}^x_{t+1|t} &= g\left(\mathcal{X}^x_{t|t}\right) + \mathcal{X}^\varepsilon_{t+1|t} \\ x_{t+1|t} &= \sum_{i=1}^{2N_a+1} W^m_i \mathcal{X}^x_{i,t+1|t} \\ P_{t+1|t} &= \sum_{i=1}^{2N_a+1} W^c_i \left[\mathcal{X}^x_{i,t+1|t} - x_{t+1|t}\right]\left[\mathcal{X}^x_{i,t+1|t} - x_{t+1|t}\right]^\top \\ \mathcal{Y}_{i,t+1|t} &= h\left(\mathcal{X}^x_{i,t+1|t}\right) + \mathcal{X}^\varepsilon_{i,t+1|t} \\ y_{t+1|t} &= \sum_{i=1}^{2N_a+1} W^m_i \mathcal{Y}_{i,t+1|t} \end{aligned}$$

3. Measurement update:

$$\begin{aligned} P^{xy}_{t+1|t} &= \sum_{i=1}^{2N_a+1} W^c_i \left[\mathcal{X}^x_{i,t+1|t} - x_{t+1|t}\right]\left[\mathcal{Y}^x_{i,t+1|t} - y_{t+1|t}\right]^\top, \\ P^{xy}_{t+1|t} &= \sum_{i=1}^{2N_a+1} W^c_i \left[\mathcal{Y}^x_{i,t+1|t} - y_{t+1|t}\right]\left[\mathcal{Y}^x_{i,t+1|t} - y_{t+1|t}\right]^\top. \end{aligned}$$

The estimate of the state vector is updated through the standard Kalman Filter recursions. The algorithm is initialized by setting the initial value to the unconditional mean and variance of the state vector.

$$\begin{aligned}
x_{0|0} &= \mathbb{E}\left[x_t\right] \\
P_{0|0} &= var\left[x_t\right] \\
x^a_{0|0} &= \left[\begin{array}{ccc} x_{0|0} & 0 & 0 \end{array}\right]^\top
\end{aligned}$$

and

$$P^a_{0|0} = \left[\begin{array}{ccc} P_{0|0} & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & R \end{array}\right].$$

F QML Standard Errors

Gallant and White (1988) show that under certain regularity conditions, the covariance matrix of the QML estimator θ^* can be estimated using the formula

$$Cov\left(\theta^*\right) = A_T^{-1}\left(\theta^*\right) B_T A_T^{-1}\left(\theta^*\right),$$

where $A_T\left(\theta^*\right)$ is the Hessian of the log-likelihood function,

$$A_T\left(\theta^*\right) = \frac{\partial^2}{\partial\theta\partial\theta'} L\left(\theta\right),$$

and B_T is a consistent estimator of the covariance matrix of the first derivative of the QML function (19). Newey and West (1987) proposed an estimator for B_T given by

$$B_T = \sum_{t=1}^{T} s_t s_t{}' + \sum_{t=1}^{L} \sum_{r=t+1}^{T} \left(1 - \frac{t}{L+1}\right)\left[s_t s_{t-r}' + s_{t-r} s_t'\right],$$

where

$$s_t = \frac{\partial}{\partial\theta} l_t\left(\theta\right).$$

G Calculation of Moments

Definition Let $X = (x_1, x_2, x_3, ..., x_N)^\top$ be normally distributed vector with mean vector μ and variance covariance matrix Σ then the moment generating function of X, denoted by $M_X(t)$ is given by:

$$M_X(t) = \mathbb{E}[\exp(X^\top t)] = \exp(\mu^\top t + \frac{t^\top \Sigma t}{2})$$

where t is an $N-$dimensional real vector.

Lemma G.1 *Let $X = (x_1, x_2, x_3, ..., x_N)^T$ be normally distributed with moment generating function, $M_X(t)$, then*

$$\frac{\partial^{q_1+...+q_N} M_X(t)}{\partial t_1^{q_1} ... \partial t_N^{q_N}} = \mathbb{E}[x_1^{q_1} ... x_N^{q_N} \exp(X^\top t)].$$

Proposition G.2 *Let $X = (x_1, x_2, x_3)^\top$ be normally distributed vector with mean vector μ and variance covariance matrix, Σ, then*

$$\begin{aligned}
\mathbb{E}(\exp(x_1)x_2) &= \exp(\mu_1 + \frac{\sigma_1^2}{2})(\sigma_{1,2} + \mu_2) \\
cov(\exp(x_1), x_2) &= \exp(\mu_1 + \frac{\sigma_1^2}{2})\sigma_{1,2} \\
cov(\exp(x_1), \exp(x_2)) &= \exp(\mu_1 + \mu_2 + \frac{\sigma_1^2 + \sigma_2^2}{2})(\exp(\sigma_{1,2}) - 1) \\
\mathbb{E}(\exp(x_1) \cdot x_2 \cdot x_3) &= \exp(\mu_1 + \frac{\sigma_1^2}{2})[(\mu_2 + \sigma_{1,2})(\mu_3 + \sigma_{1,3}) + \sigma_{2,3}] \\
cov(\exp(x_1) \cdot x_2, x_3) &= \exp(\mu_1 + \frac{\sigma_1^2}{2})[\sigma_{1,3}(\mu_2 + \sigma_{1,2}) + \sigma_{2,3}]
\end{aligned}$$

Proof The proof follows directly from applying the previous lemma.

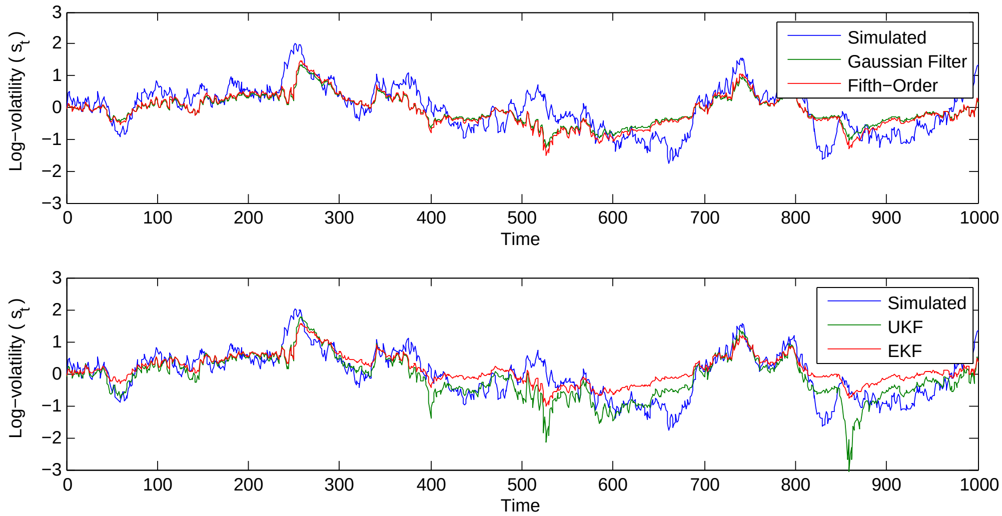

FIGURE 1

Sensitivity: This figure compares a simulated time series of 1000 observations for the standard Stochastic Volatility Model with its filtered estimates. The filtered estimates in top figure were calculated using a fifth order approximation and an infinite order of approximation (Gaussian filters). The parameter values used for the simulation as well as for the filtered estimates are $\phi = 0.98$, $\sigma_\varepsilon = 0.1414$, $\overline{\sigma} = 1$ and $\rho = -0.5$.

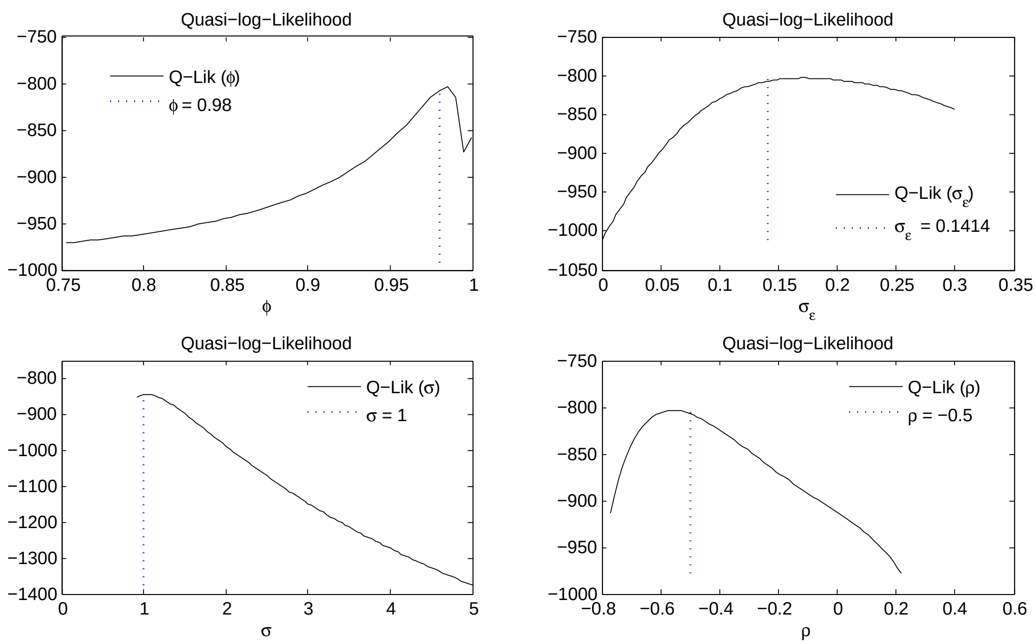

FIGURE 2

Sensitivity: This figure plots the quasi-likelihood function of a the standard Stochastic Volatility Model for different sets of parameter vectors. The plots show the quasi-log-likelihood function of the data for different values of ϕ (top left), σ_ε (top right), $\overline{\sigma}$ (bottom right) and ρ (bottom left). The vertical dashed lines represent the parameter values that were used to simulate the data.

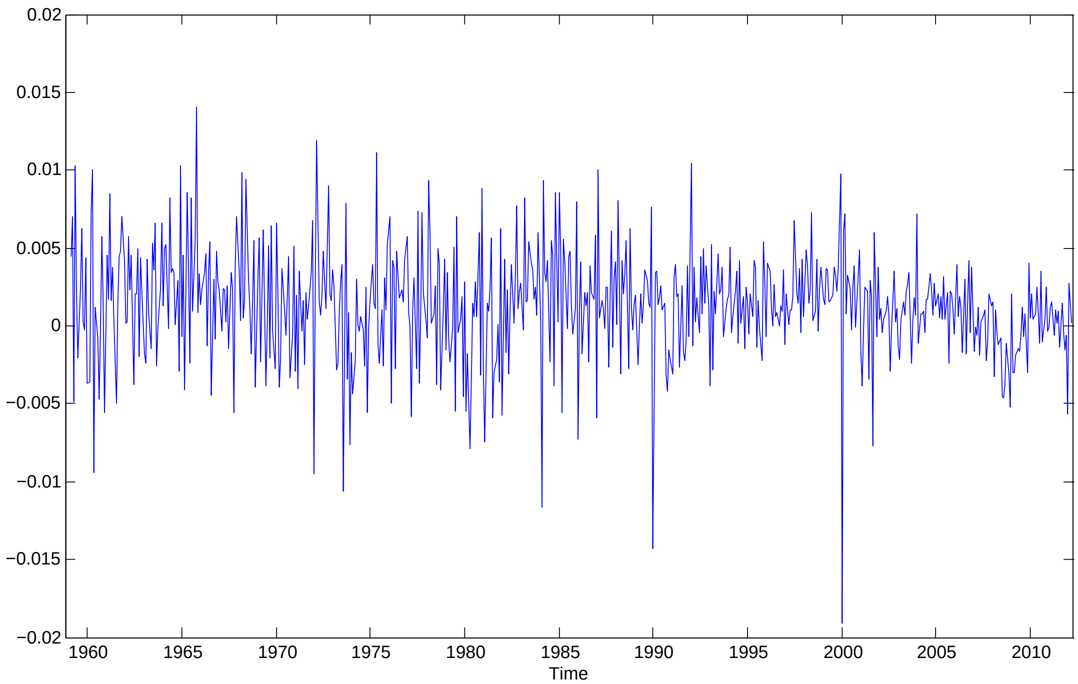

FIGURE 3

Returns: This figure plots the monthly log-consumption growth on the monthly real consumption series per capita for nondurables and services from January 1959 to March 2012.

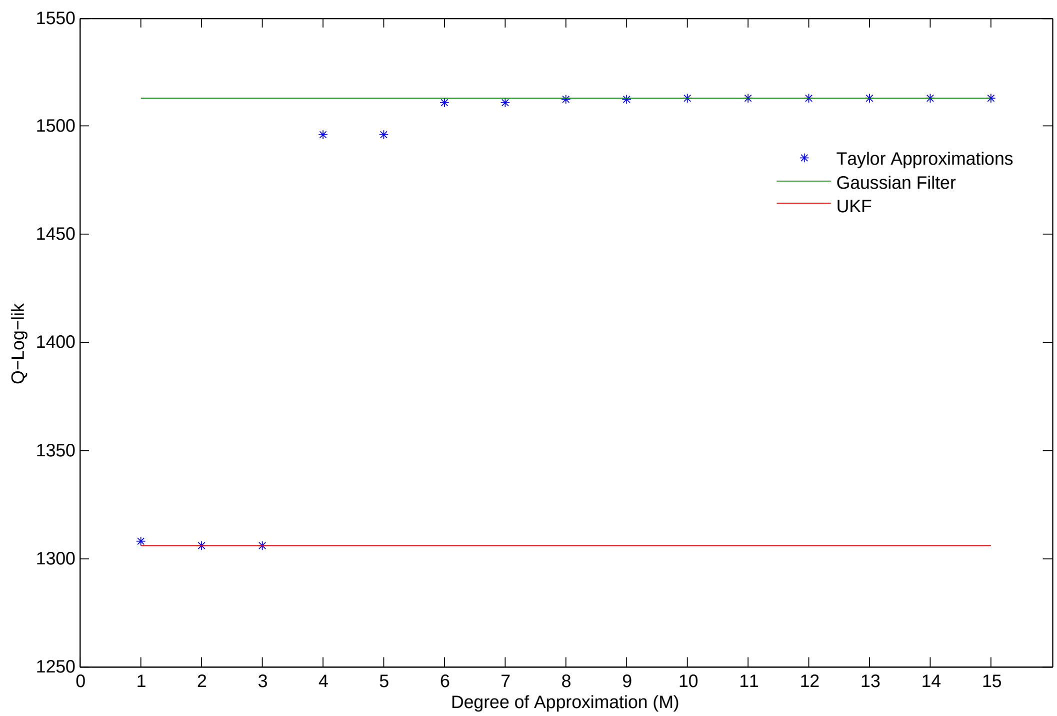

FIGURE 4

Order of Approximation: This figure plots the quasi-likelihood function of the Model by Brandt and Kang (2004). The plot compares the quasi-likelihood function for different orders of approximation, $M = 1, 2, ..., 15$ (asterisks), with quasi-likelihood functions constructed with the Unscented Kalman Filter and an infinite order of approximation filter or Gaussian filters (continuous lines).

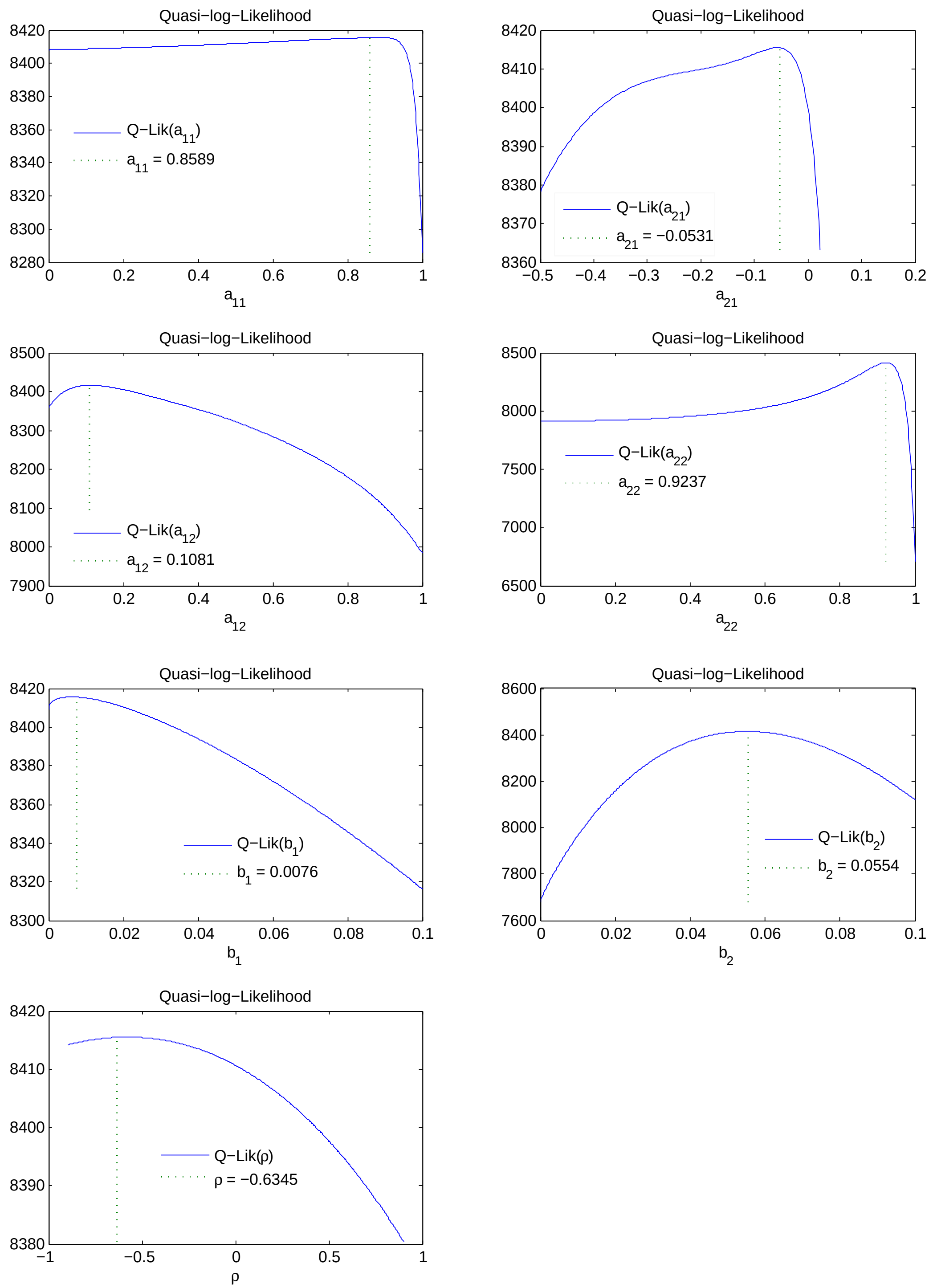

Figure 5

Sensitivity: This figure plots the quasi-likelihood function of the Model by Brandt and Kang (2004). The quasi-likelihood function is based on a random draw of $T = 5000$ returns simulated from the model with the parameter values .

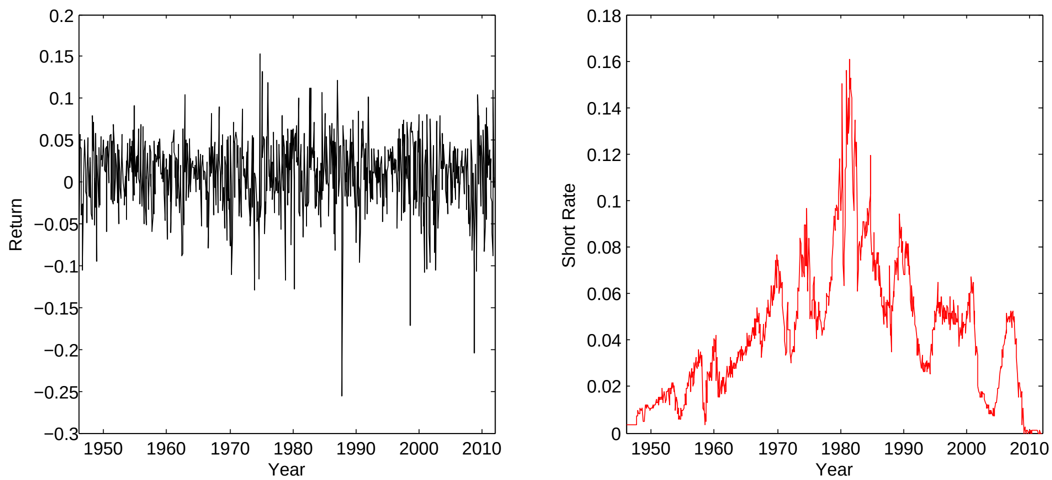

FIGURE 6

Returns: This figure plots the monthly returns on the value weighted CRSP index as well as the short rate from January 1946 through December 2011.

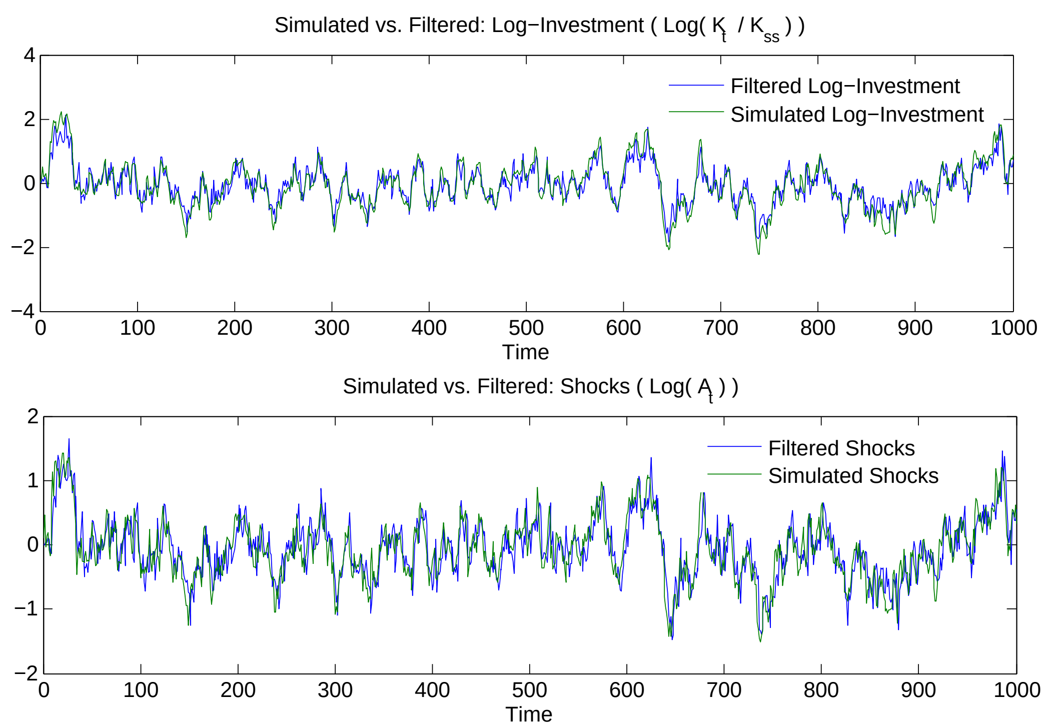

FIGURE 7

Filtering: This figure plots the filtered estimates of the state variables evaluated in a simulated sample path of size $T = 1000$ with parameter values $\beta = 0.95$, $\delta = 0.15$, $\alpha = 0.30$, $\rho = 0.90$,$\gamma = 3$,$\sigma_y = 0.30$ and $\sigma_\varepsilon = 0.2$

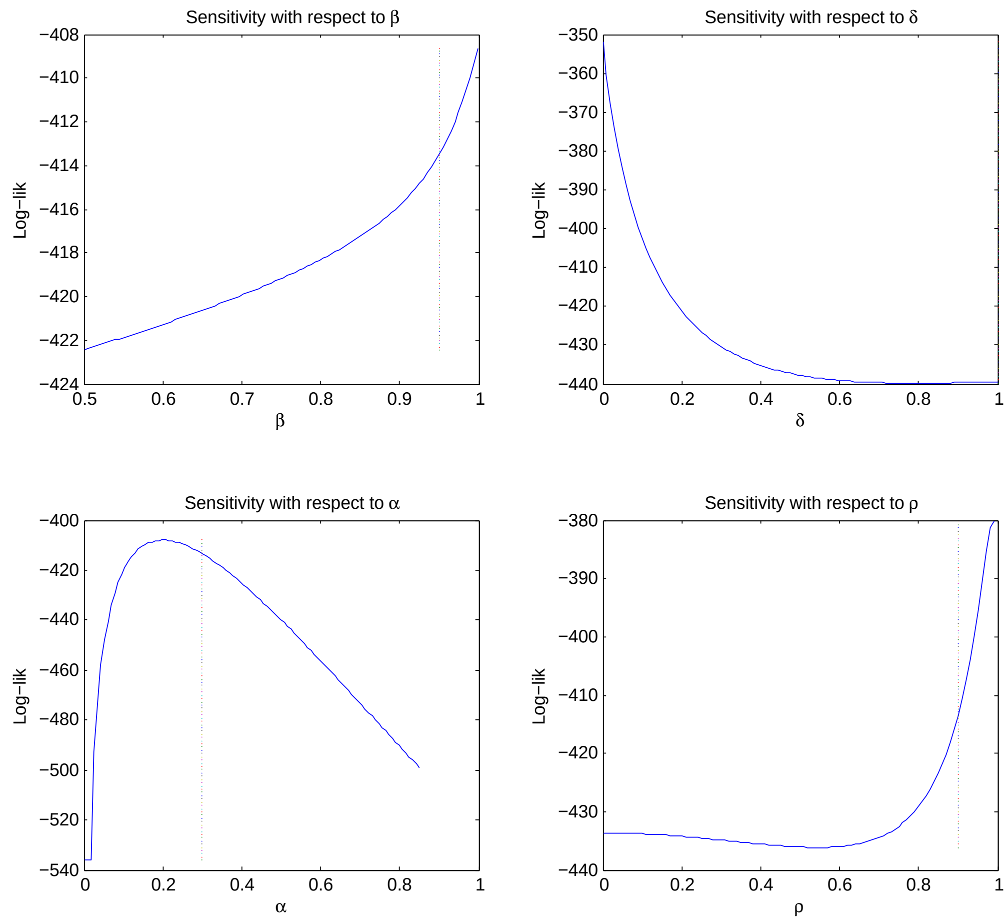

FIGURE 8

Sensitivity: This figure plots the likelihood function evaluated at a sample path of size $T = 500$ with parameter values $\beta = 0.95$, $\delta = 0.15$, $\alpha = 0.30$, $\rho = 0.90$,$\gamma = 3$,$\sigma_y = 0.30$ and $\sigma_\varepsilon = 0.2$.

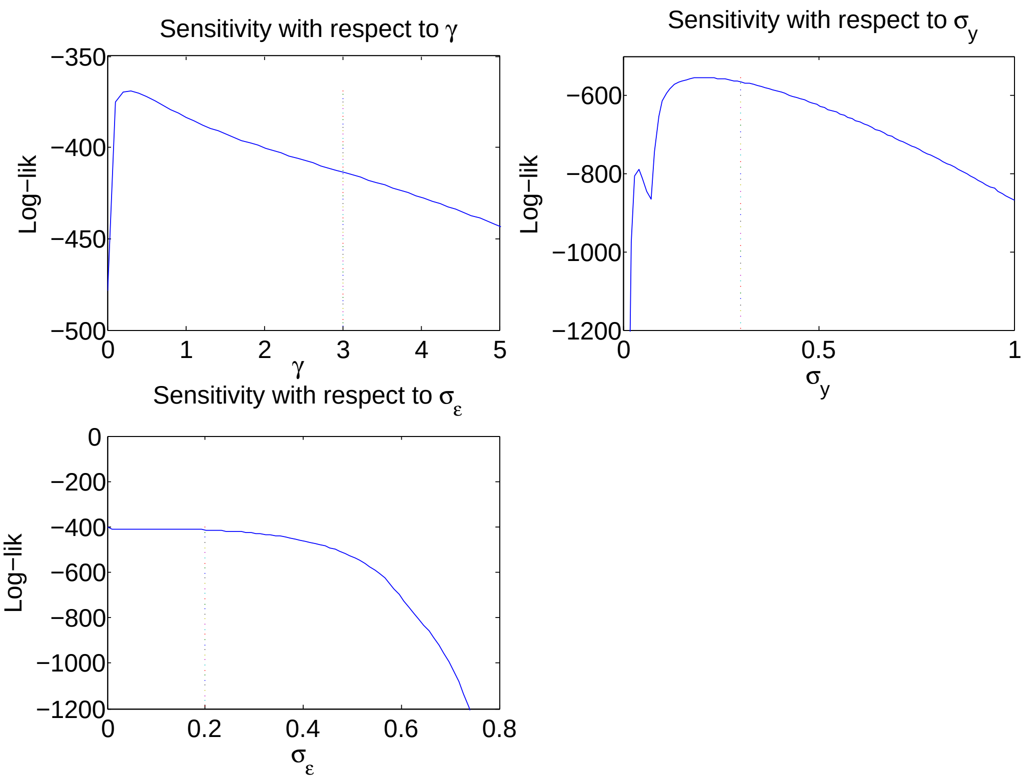

FIGURE 9

Sensitivity: This figure plots the likelihood function evaluated at a sample path of size $T = 500$ using the parameter values $\beta = 0.95$, $\delta = 0.15$, $\alpha = 0.30$, $\rho = 0.90$,$\gamma = 3$,$\sigma_y = 0.30$ and $\sigma_\varepsilon = 0.2$

H Figures and Tables

Table 1

Stochastic Volatility Model: Simulation Results

Filtering Method	Mean Squared Error		Mean Squared Error	
	Average	Std. Dev.	Average	Std. Dev.
EKF	0.3795	0.1343	2.5508	1.4006
UKF	0.3067	0.1103	2.5689	1.4368
TKF 2	0.3043	0.0944	2.5424	1.3914
TKF 3	0.3043	0.0944	2.5424	1.3914
TKF 4	0.3037	0.0888	2.5416	1.3897
TKF 5	0.3037	0.0888	2.5416	1.3897
TKF 6	0.3105	0.0912	2.5417	1.3900
TKF 7	0.3105	0.0912	2.5417	1.3900
TKF 8	0.3123	0.0918	2.5417	1.3901
TKF 9	0.3123	0.0918	2.5417	1.3901
TKF 10	0.3127	0.0919	2.5417	1.3901
TKF 11	0.3127	0.0919	2.5417	1.3901
Gaussian	0.3127	0.0920	2.5417	1.3901

State estimation results. This table presents the mean and variance of the mean squared errors (MSE) of the model:

$$\begin{aligned} y_t &= \eta_t \cdot \sigma_t \\ \log \sigma_t^2 &= \log \overline{\sigma}^2 + \phi \log \sigma_{t-1}^2 + \varepsilon_t,\ \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right), \end{aligned}$$

The results are based on 250 independent samples of $T = 1000$ simulated from the model with the parameters $\phi = 0.98$, $\sigma_\varepsilon = 0.1414$, $\overline{\sigma} = 1$ and $\rho = -0.5$.

Table 2

Stochastic Volatility Model: Simulation Results

Method	ϕ	σ_ε	$\overline{\sigma}$	ρ
Parameter Value	0.9800	0.1414	1.0000	-0.5000
EKF	0.9348 (0.1454)	0.1491 (0.0701)	1.2402 (0.4162)	-0.3598 (0.4479)
UKF	0.9287 (0.1258)	0.1614 (0.1050)	1.2926 (0.4539)	-0.4383 (0.4656)
TKF3	0.9241 (0.1698)	0.1601 (0.1012)	1.2926 (0.4277)	-0.4805 (0.4663)
TKF5	0.9292 (0.2085)	0.1753 (0.1030)	0.966 (0.3975)	-0.4663 (0.4353)
TKF7	0.9564 (0.0693)	0.1781 (0.1035)	0.8502 (0.4103)	-0.4852 (0.4091)
TKF9	0.9405 (0.1629)	0.1907 (0.1013)	0.8358 (0.5483)	-0.4418 (0.3999)
GF	0.955 (0.0820)	0.1542 (0.0899)	0.9147 (0.4263)	-0.5393 (0.3840)

Finite sample properties of the Quasi-Maximum Likelihood estimator. This table presents the sample mean and standard deviation of the Quasi-Maximum Likelihood estimates of the model:

$$\begin{aligned} y_t &= \eta_t \cdot \sigma_t \\ \log \sigma_t^2 &= \log \overline{\sigma}^2 + \phi \log \sigma_{t-1}^2 + \varepsilon_t,\ \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right), \end{aligned}$$

The results are based on 250 independent samples of $T = 500$ simulated from the model with parameters in the first row.

Table 3
Data: Descriptive Statistics

Monthly Consumption Growth	
Mean	0.0013
Std. Dev.	0.0036
Max	0.0140
Min	-0.0191
Median	0.0014
Skewness	-0.4148
Kurtosis	5.6120
Autocorrelation	
1-month	-0.1869
6-month	0.0688
12-month	-0.0242
24-month	-0.0955

This table presents descriptive statistics of monthly log-consumption growth on the monthly real consumption series per capita for nondurables and services from January 1959 to March 2012.

Table 4

Parameter Estimates

	ϕ	σ_ε	$\overline{\sigma}$	ρ	μ_C	
EKF	0.9575 (0.0862)	0.2997 (0.8654)	0.1591 (0.2196)	0.9991 (2.5046)	0.0014 (0.0002)	-586.08
UKF	0.9623 (0.0616)	0.0198 (0.0508)	0.0035 (0.0012)	0.9983 (0.0004)	0.0012 (0.0027)	2696.67
TKF 3	0.9623 (0.0287)	0.0221 (0.0218)	0.0035 (0.0007)	0.8923 (0.0005)	0.0012 (0.0012)	2696.67
TKF 5	0.9626 (0.0724)	0.0266 (0.0465)	0.0035 (0.0006)	0.7436 (0.0007)	0.0012 (0.0014)	2696.71
TKF 7	0.9627 (0.0203)	0.0216 (0.0332)	0.0035 (0.0006)	0.9179 (0.0045)	0.0012 (0.0018)	2696.71
TKF 9	0.9627 (0.0446)	0.0228 (0.0297)	0.0035 (0.0005)	0.8683 (0.0003)	0.0012 (0.0011)	2696.71
TKF 11	0.9626 (0.0204)	0.0336 (0.0146)	0.0035 (0.0002)	0.5867 (0.0008)	0.0012 (0.0002)	2696.71
GF	0.9626 (0.0205)	0.0353 (0.0160)	0.0035 (0.0001)	0.5574 (0.0006)	0.0012 (0.0002)	2696.71

This table describes presents the the Quasi-maximum likelihood estimates of the model:

$$\begin{aligned} \Delta \ln(C_{t+1}) - \mu_c &= \overline{\sigma}\eta_{t+1} \\ s_{t+1} &= \phi s_t + \varepsilon_{t+1},\ \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right), \end{aligned}$$

The estimates are for monthly real consumption growth on the monthly vintage series from the Federal Reserve Bank of Philadelphia from January 1959 to March 2012. Each row contains the estimates under the filtering techniques based on different orders of approximation. Standard errors are reported between parenthesis.

Table 5

Finite sample properties of Quasi-Maximum-Likelihood estimator.

Parameters	Model A			Model B		
	True Value	Average	Std. Dev.	True Value	Average	Std. Dev.
a_{11}	0.8589	0.9111	0.1158	0.8313	0.8277	0.2100
a_{21}	-0.0529	-0.0099	0.0983	-0.0211	-0.0377	0.2350
a_{12}	0.1084	0.3474	0.3015	0.1168	0.3771	0.3617
a_{22}	0.9226	0.8792	0.1273	0.9110	0.8181	0.2125
b_{11}	0.0076	0.0033	0.0070	0.0064	0.0031	0.0081
b_{22}	0.0553	0.0347	0.0375	0.0561	0.0812	0.01123
ρ	-0.6336	0.1687	0.8037	-0.4577	-0.0018	0.5760
$\overline{\mu}$	0.0067	0.0067	0.0016	0.0065	0.0067	0.0015
$\overline{\sigma}$	0.0418	0.0523	0.0046	0.0385	0.0524	0.0045
ρ_μ	-	-	-	-0.0866	0.1154	0.7475
ρ_σ	-	-	-	-	-	-

Parameters	Model C			Model D		
	True Value	Average	Std. Dev.	True Value	Average	Std. Dev.
a_{11}	0.8658	0.7841	0.2581	0.8677	0.8037	0.2501
a_{21}	-0.0885	-0.0259	0.2425	-0.1292	-0.1065	0.2374
a_{12}	0.0861	0.3229	0.2872	0.0947	0.3121	0.3052
a_{22}	0.8973	0.8727	0.1687	0.9086	0.8540	0.1880
b_{11}	0.0060	0.0065	0.0078	0.0047	0.0044	0.0069
b_{22}	0.0614	0.0104	0.0268	0.0591	0.0338	0.0755
ρ	-0.5584	-0.0211	0.5872	-0.5621	0.0036	0.6259
$\overline{\mu}$	0.0062	0.0063	0.0014	0.0062	0.0063	0.0014
$\overline{\sigma}$	0.0382	0.0506	0.0035	0.0382	0.0508	0.0034
ρ_μ	-	-	-	-0.0517	0.1438	0.7477
ρ_σ	-0.2541	-0.6629	0.3716	-0.2430	-0.5841	0.4675

This table describes the sampling distribution of the Quasi-maximum likelihood of the model:

$$y_t = \overline{\mu}\exp(x_{1t}) + \overline{\sigma}\exp(x_{2t})\,x_{3t},$$

and

$$x_t = \widetilde{A}x_{t-1} + \underline{w_t} \text{ with } \underline{w_t} \sim N\left(0, \widetilde{\Sigma}\right),$$

where

$$\widetilde{A} = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \widetilde{\Sigma} = \begin{bmatrix} b_{11} & \rho\sqrt{b_{11}b_{22}} & \rho_\mu\sqrt{b_{11}} \\ \rho\sqrt{b_{11}b_{22}} & b_{22} & \rho_\sigma\sqrt{b_{22}} \\ \rho_\mu\sqrt{b_{11}} & \rho_\sigma\sqrt{b_{22}} & 1 \end{bmatrix}.$$

The results are based on 500 independent samples of $T = 792$ returns simulated from the model with the parameters in the first column.

Table 6

Data: Descriptive Statistics

	Market Index	Short rate
Mean	0.0083	0.0036
Std. dev.	0.0435	0.0025
Max	0.1532	0.0134
Min	-0.2554	0.0000
Median	0.0127	0.0034
Skewness	-0.7680	0.9463
Kurtosis	5.6443	4.2273
Autocorrelation		
1-month	0.0908	0.9684
6-month	-0.0556	0.8907
12-month	0.0348	0.8080
24-month	-0.0008	0.6327

This table presents descriptive statistics of monthly log-returns on the value-weighted CRSP index and the short rate from January 1946 to December 2011. The short rate is the yield on a one-month Treasury bill.

TABLE 7

Parameter Estimates

Parameters	Model A		Model B		Model C		Model D	
	Estimate	**Std. Error**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
a_{11}	0.9436	0.0662	0.9586	0.0161	0.9999	0.0355	0.9927	0.0859
a_{21}	-0.0778	0.1146	-0.0285	0.0065	-0.0202	0.0232	-0.0135	0.0000
a_{12}	0.3799	0.1171	0.3054	0.0253	0.2873	0.0261	0.6779	0.0067
a_{22}	0.7317	0.1298	0.8750	0.0001	0.8576	0.0495	0.8523	0.0464
b_{11}	0.1684	0.0296	0.1739	0.0015	0.0474	0.0762	0.0615	0.0074
b_{22}	0.0002	0.0316	0.0024	0.0015	0.0054	0.0019	0.0053	0.0277
ρ	-0.1306	0.0091	-0.9000	0.0014	-0.8999	0.0133	-0.8897	0.0034
$\overline{\mu}$	0.0047	0.0165	0.0047	0.0012	0.0046	0.0011	0.0047	0.0035
$\overline{\sigma}$	0.0434	0.0070	0.0437	0.0002	0.0440	0.0211	0.0436	0.0002
ρ_μ	-	-	-0.6642	0.0170	-	-	-0.1902	0.0135
ρ_σ	-	-	-	-	-0.9000	0.0002	-0.8942	0.0203
Q-lik	1366.23		1373.35		1388.43		1395.50	

This table describes presents the the Quasi-maximum likelihood estimates of the model:

$$y_t = \overline{\mu} \exp(x_{1t}) + \overline{\sigma} \exp(x_{2t})\, x_{3t},$$

and

$$x_t = \widetilde{A} x_{t-1} + \underline{w_t} \text{ with } \underline{w_t} \sim N\left(0, \widetilde{\Sigma}\right),$$

where

$$\widetilde{A} = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \widetilde{\Sigma} = \begin{bmatrix} b_{11} & \rho\sqrt{b_{11}b_{22}} & \rho_\mu\sqrt{b_{11}} \\ \rho\sqrt{b_{11}b_{22}} & b_{22} & \rho_\sigma\sqrt{b_{22}} \\ \rho_\mu\sqrt{b_{11}} & \rho_\sigma\sqrt{b_{22}} & 1 \end{bmatrix}.$$

The estimates are for quarterly returns on the value-weighted CRSP index in excess of the one-month Treasury bill from the January 1953 to December 2011.