# Filtering in Asset Pricing

by

Alberto Romero

B.Sc., ITAM, 2002

M.Math. University of Waterloo, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

**DOCTOR OF PHILOSOPHY**

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Business Administration)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

September 2013

# Abstract

In the first chapter of this thesis, I propose a nonlinear filtering method to estimate latent processes based on the Taylor series approximations. The filter extends conventional methods such as the extended Kalman filter or the unscented Kalman filter and provides a tractable way to estimate filters of any order. I apply the filter to different models and demonstrate that this method is a good approach for the estimation of unobservable states as well as for parameter inference. I also find that filters with Taylor approximations can be as accurate as conventional Monte Carlo filters and computationally more efficient. Through this chapter I show that filters with Taylor approximations are a good approach for a number of problems in finance and economics that involve nonlinear dynamic modeling.

In the second chapter, I investigate the recently documented, large time-series variation in the empirical market Sharpe ratio. I revisit the empirical evidence and ask whether estimates of Sharpe ratio volatility may be biased due to the limitations of the standard ordinary least squares (OLS) methods used in estimation. Based on simulated data from a standard calibration of the long-run risks model, I find that OLS methods used in prior literature produce Sharpe ratio volatility five times larger than its true variability. The difference arises due to measurement error. To address this issue, I propose the use of filtering techniques that account for the Sharpe ratio's time variation. I find that these techniques produce Sharpe ratio volatility estimates of less than 15% on a quarterly basis, which match more closely the predictions of standard asset pricing models.

# Preface

This dissertation is original, unpublished, independent work by the author, Alberto Romero.

# Table of Contents

# List of Tables

# List of Figures

# Acknowledgments

Completing my degree at the Sauder School of Business would not have been possible without the continued support from a great number of faculty, colleagues, friends, and family during these six years.

I owe an infinite amount of gratitude to my Ph.D. supervisors Adlai Fisher and Lorenzo Garlappi, who provided guidance and support on so many levels, but foremost as great mentors.

I want to thank Harjoat Bhamra, Murray Carlson, Nando de Freitas, Howard Kung and Jason Chen for their help, encouragement and guidance on my research endeavors. Jan Bena, Ron Giammarino, Alan Kraus, Kai Li, Hernan Ortiz-Molina and Tan Wang provided much welcome advice throughout my studies.

My colleagues and friends Milka Dimitrova, Vincent Gregoire, Alejandra Medina and Gonzalo Morales have been helpful every step of the way.

My parents, Maria del Carmen and Jose Humberto, never gave up on me. My parents-in-law, Luz Maria and Modesto, have been a source of support and encouragement.

Special thanks go to Zara Contractor and Daniella Santos Coy for their help and support during the last couple of years during the Ph.D. program.

Finally, it was my wife Ana that made my time in Vancouver truly special. Without her constant support I could not have completed my thesis. No words could describe how grateful I am to her for always believing in me, and keeping up with the ups and downs of this journey. Lastly, I want to thank my son David, whose birth gave me the motivation I needed to complete my Ph.D.

Thanks to all of you and to all that I have forgotten to mention by name.

# Dedication

*To Ana and David.*

# Chapter 1

# Introduction

One of the most important research topics in financial economics is the impact of noisy information on investment decisions. As a result, the use of latent variables in economic models has become crucial, since those variables capture unobserved changes in the economic environment. Therefore, developing statistical methods for estimating latent variables from information observed with noise is of utmost importance. A standard approach for estimating these variables is the use of filtering methods. *Filtering*, in general, refers to an extraction process, and *statistical filtering* refers to an algorithm for extracting a latent state variable from noisy measurements. In this thesis I develop a statistical technique for estimating unobserved state variables and explore the use of filtering methods in asset pricing.

In the first chapter of this thesis, "Filtering via Taylor Series Approximations," I propose a nonlinear filtering method to estimate latent processes based on Taylor series approximations. The method can be applied to both state and parameter inferences and generalizes conventional methods such as the extended Kalman filter (EKF) or the unscented Kalman filter (UKF). My findings show that filters with Taylor approximations can be as accurate as standard particle filters for state estimation. Based on different empirical applications, I provide evidence that my filtering method is a good approach for econometric inference of dynamic models. The estimation technique I propose can be applied to a number of empirical and theoretical problems that involve calculating conditional expectations based on noisy information.

The second chapter of this thesis, "On the Volatility of the Market Sharpe Ratio," is based on recent literature that empirically documents large time-series variation in the market Sharpe ratio, which has spurred theoretical explanations for this phenomenon. I revisit the empirical evidence and ask whether estimates of Sharpe ratio volatility might be biased due to limitations of the standard ordinary least squares (OLS) methods used in estimation. Based on simulated data from a standard calibration of the long-run risks model, I find that OLS methods used in the literature produce Sharpe ratio volatility five times larger than its true variability. The difference arises due to measurement error. To address this issue I propose the use of filtering

techniques that account for the Sharpe ratio's time variation. I find that these techniques produce Sharpe ratio volatility estimates that match more closely the predictions of standard asset pricing models. Additionally, my results have practical implications for portfolio allocation, where upward-biased estimates of Sharpe ratio volatility imply excessive portfolio rebalancing.

# Chapter 2

# Filtering via Taylor Series Approximations

## 2.1 Introduction

Filtering is a statistical tool that recovers unobservable state variables using measurements that are observed with *noise*.[1] 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 in which 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. The problem is that the solution to the filtering problem is known analytically only in a limited number of settings, such as the linear, and alternative solutions are required.

In this chapter, I propose a nonlinear filtering method to estimate unobserved state variables which is based on an efficient calculation of Taylor approximations. The method can be applied to both state and parameter inferences and generalizes conventional methods such as the EKF or the UKF. My findings show that filters with Taylor approximations can be as accurate as standard particle filters for state estimation. The importance of the filter with Taylor series is that it overcomes a number of difficulties previously documented in the filtering literature (Fernández-Villaverde and Rubio-Ramírez, 2007; Fernández-Villaverde, Rubio-Ramírez, and Santos, 2006). First, it allows for arbitrary nonlinearities in the data-generating process. Second, the filtering calculations are as efficient as the standard Kalman filter because only function evaluations are required to calculate the recursions. Third, the order of approximation of the

[1] This technique has been the subject of considerable research during the past decades due to its numerous applications in science and engineering, such as satellite navigation systems, tumor identification and weather forecasting.

filter can be chosen exogenously by the researcher. Fourth, in addition to state estimation, the filter with Taylor series can be applied for inference purposes since a quasi-likelihood function is obtained in the filtering recursions and quasi-likelihood methods can be applied (Bollerslev and Wooldridge, 1992; White, 1982). In addition, the quasi-likelihood functions of the filter with Taylor series are continuous with respect to the parameter values, and conventional methods for numerical optimization can be applied to conduct statistical inference. Lastly, the filter is flexible enough to include several nonlinear observation equations to improve the state and parameter estimation.

The filter approximates all the densities involved in the state estimation process with Gaussian distributions. The theoretical foundation for these approximations is that any probability density function can be approximated as closely as desired by the sum of Gaussian density functions (Ito and Xiong, 2000; Maz'ya and Schmidt, 1996), where the first and second moments are necessary to characterize the whole distribution. In addition, the filter can be combined with Monte Carlo simulations to handle a more general class of models that involve discrete state variables via the Rao-Blackwellised particle filter (Doucet, De Freitas, Murphy, and Russell, 2000).

I test the proposed filter in a number of nonlinear models that involve latent variables previously studied in the finance, economics and filtering literature. The first application is the stochastic volatility model (Andersen, Bollerslev, Diebold, and Ebens, 2001; Andersen, Bollerslev, Diebold, and Labys, 2003; Andersen and Sørensen, 1996; Broto and Ruiz, 2004). I start by studying the performance of the filter with simulated data. The filter with Taylor series generates volatility estimates as accurate as those of the particle filters and at least four times faster. I also find that quasi-maximum likelihood methods are a good approach for parameter estimation. My simulation exercises suggest that the filtering method with Taylor series approximations is an alternative approach for both state and parameter inference. Finally, I estimate the parameters, for different orders of approximation, of an endowment process with stochastic volatility using a series of US data for monthly consumption growth. For higher approximation orders, I found evidence of stochastic volatility comparable with the recent findings by Bidder and Smith (2011) and Ludvigson (2012) which suggest that the stochastic volatility model is a good representation for consumption growth, as posited 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) and recently used by Boguth, Carlson, Fisher, and Simutin (2011) in the conditional asset pricing literature. In this setup, the conditional mean and volatility of stock returns are modeled 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 the study of the contemporaneous and intertemporal relationships between expected returns and risk. In this case, by including an additional observation equation I find that the filters based on

Taylor series generate estimates of expected returns and volatilities as accurately as the particle filters do.

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 (Judd, 1998; Schmitt-Grohe and Uribe, 2004) 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 these models since a quasi-likelihood function can be constructed instead of Monte Carlo simulation methods.

As a robustness check, I test the filter in a high-dimensional multivariate stochastic volatility model and a standard highly nonlinear model from the filtering literature. I find that the filter with Taylor series provides accurate state estimates that are comparable with those obtained with particle filters. More importantly, I confirm that the filter with Taylor series is computationally more efficient than standard particle filters.

### 2.1.1 Related Literature

A number of applications involve nonlinear VAR processes. These nonlinearities complicate the filtering process as well as the parameter inference procedures because the Kalman filter is no longer an optimal solution. To resolve this issue, different lines of research have emerged. One strand of research is based on deterministic filtering and uses deterministic recursions to compute the mean and variance of the state variables given the observed information. Two widely used algorithms have been successfully applied: the EKF (Jazwinski, 1970) and the UKF (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.[2] A recent extension of this approach is the Smolyak Kalman filter, proposed by Winschel and Krätzig (2010), which extends the UKF by applying the Smolyak quadratures to construct the filtering recursions. This chapter extends these approaches by allowing an arbitrary order of approximation. The filter is based on the efficient Taylor series expansions recently used in Savits (2006) and Garlappi and Skoulakis (2010) in two manners: 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 filters with Taylor series approximations proposed in this paper fall into the deterministic filtering literature and extend the current techniques for nonlinear filtering.

The second line of research for nonlinear filtering is based on Monte Carlo techniques. These filtering techniques, also called particle filtering techniques, are based on Monte Carlo

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

simulation with sequential importance sampling. The overall goal is to directly implement optimal Bayesian estimation by recursively approximating the complete posterior state density through Monte Carlo methods (Gordon, Salmond, and Smith, 1993; Pitt and Shephard, 1999). Different extensions of the particle filter have been proposed in the filtering literature, such as the Rao-Blackwellised particle filter (Doucet, De Freitas, Murphy, and Russell, 2000) and the unscented particle filter (Van Der Merwe, Doucet, De Freitas, and Wan, 2001). The current approach to evaluating the likelihood of a nonlinear state-space model is dominated by particle filters; its extensions are described in Doucet, de Freitas, Gordon, and Smith (2001). Particle filters are often an alternative to the EKF or UKF and have the advantage that, with sufficient samples, they approach the Bayesian optimal estimate, improving the accuracy of the EKF or UKF. However, when the simulated sample is not sufficiently large, particle filters might suffer from sample impoverishment. In this paper, I show that the filter with Taylor series can provide state estimates as accurate as those obtained by standard particle filters by including additional observation equations, such as the squared or cubed observation equations of a standard model. By adding observation equations to the model, we can also achieve a better parameter identification.

Tanizaki and Mariano (1996) are the first to suggest using Taylor series approximations to resolve the filtering problem as well as the biases that arise while taking first- and second-order approximations. However, they propose the use of Monte Carlo simulations to avoid numerical integration. Instead, I apply Taylor series approximations in filtering and use the efficient recursions of Savits (2006) to estimate the derivatives of a function with several variables as an argument.

The rest of this chapter is structured as follows: Section 2.2 outlines the general filtering problem, Section 2.3 presents the filtering techniques with Taylor series and Section 2.4 describes the quasi-maximum likelihood approach for parameter estimation in filtering. Section 2.5 presents three different applications and describes the data, results and empirical findings; Section 2.6 provides a set of robustness checks and Section 2.7 concludes the chapter.

## 2.2 Nonlinear Filtering

State-space models are mathematical tools commonly used to represent dynamic systems that involve unobserved state variables.[3] A state-space representation is characterized by a set of measurements and a state transition, usually obtained from a theoretical 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 $y_t$ be a $p$-dimensional vector of observables. It is generally assumed that the states of the system follow a first-order Markov

[3]See Hamilton (1994) and Kim and Nelson (1999) for a standard introduction to state-space models.

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 the transition and measurement equations, which are expressed as follows:

$$\text{Observation Equation:} \qquad y_t = h\left(x_t\right) + v_t, \tag{2.1}$$

$$\text{Transition Equation:} \qquad x_{t+1} = g\left(x_t\right) + \varepsilon_{t+1}, \tag{2.2}$$

where $v_t$ and $\varepsilon_t$ are $p$-dimensional and $N$-dimensional distributed noise vectors with variance–covariance matrices $R$ and $Q$, respectively. In this case, Eq. (2.1) represents the observation equation, while Eq. (2.2) represents the transition equation. Intuitively the function $h$ defines the measurement based on the current state and the function $g$ characterizes the current state from the previous state. The mappings $h : \mathbb{R}^N \longrightarrow \mathbb{R}^p$ and $g : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ are assumed to be continuous and smooth.

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 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, the mathematical object that is usually analyzed is the marginal distribution, or marginal density of the state conditional on the set of observations available, which is denoted by $p(x_t|y_1, ..., y_t)$. If $h$ and $g$ are linear, then Eqs. (2.1) and (2.2) define a linear filtering problem; moreover, if $v_t$ and $\varepsilon_t$ are normally distributed, then the filtering problem has a well-known solution given by the Kalman filter (Kalman, 1960). In the linear case, the conditional density, $p(x_t|y_1, ..., y_t)$, is Gaussian with mean and variance constructed recursively based on the set of observations and the state-space representation. If either of the mappings $h$ or $g$ is nonlinear, then the filtering problem is nonlinear and no standard solution exists.

## 2.3 Filtering Based on Taylor Series Expansions

A number of solutions have been proposed to solve the nonlinear filtering problem.[4] If the nonlinear models can be expressed in a state-space setting, then the Kalman filter may be useful by calculating linearizations at each time step, so that the standard filter recursions can be applied. This approach is known as the extended Kalman filter (EKF).[5] The EKF reverts to optimal Kalman filters when the problems become linear. As a result, the EKF can

[4] An extensive review of nonlinear filtering from a theoretical and empirical perspective is provided by Crisan and Rozovskii (2011).

[5] See Appendix A.4 for a detailed explanation of the EKF.

yield approximate minimum-variance "try at your own risk" category. Indeed, Anderson and Moore (1979) caution that the EKF "can be satisfactory on occasions." Moreover, Julier and Uhlmann (1997) document a number of biases generated by the EKF, and as a result propose an improvement which is the unscented Kalman filter (UKF).[6]

The UKF relies on the idea that approximating the moments of a transformed random variable is simpler than approximating the density function itself. 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 integrands to be analyzed. As shown by Julier and Uhlmann (1997), this approximation is comparable to a second-order Taylor approximation of the state and observation equations. Winschel and Krätzig (2010) find that the UKF is an attempt to solve the *curse of dimensionality* generated by the number of integrands in the filtering recursions; however, the filter generates another curse in terms of approximation errors. As the dimension of the problem increases, the number of points used by the UKF rises linearly. Unfortunately, the accuracy of the numerical integration decreases with the dimensionality and nonlinearity of the integrands. Therefore, this curse of approximation errors has an effect in state and parameter estimation. The UKF is therefore restricted to tractable nonlinearities (such as a low-order polynomial) and a small number of states.

To overcome this issue, I propose the use of higher Taylor series expansions for nonlinear filtering. This technique assumes that the nonlinear functions that define the state-space have a Taylor series expansion. In order to calculate the moments involved in the filtering recursions, the filter uses the moment calculations of a Taylor series with a level of approximation previously chosen by the researcher. The moments of the Taylor series expansion are then used in the standard Kalman filter recursions. This approach extends the EKF and UKF to any order of approximation, and its computational efficiency is comparable to that of the standard Kalman filter. The use of Gaussian distributions to approximate filters is the basis of deterministic filtering algorithms. These techniques have been analyzed by Ito and Xiong (2000) in the filtering literature. Additionally, the filters with Taylor series can be applied for parameter inference via quasi-maximum likelihood methods, first introduced by White (1982) and analyzed in Bollerslev and Wooldridge (1992).

The following sections present the filtering method based on Taylor approximations. I first introduce the use of Gaussian densities in nonlinear filtering and explain how the standard Kalman filter is applied to estimate the first two moments of the unobserved state variables. The Kalman filter makes use of means, variances and covariances of nonlinear transformations of normally distributed random vectors. I show how to estimate these moments with Taylor series approximations and present theoretical results that help estimating them efficiently. At

[6]See Appendix A.5 for a detailed explanation of the UKF.

the end of the section, I summarize the results in a general algorithm, and, finally, I discuss how to apply the filters with Taylor approximations to a more general class of state variables via standard Rao-Blackwellisation methods.

### 2.3.1 Gaussian Densities for Filtering

This section introduces Gaussian densities for nonlinear filtering and describes how the mean and variance of the unobserved state variables are approximated with the standard Kalman filter. This approximation requires calculating expected values of nonlinear transformations of normally distributed random vectors which may not have a closed form. I describe how the Taylor series can be applied to estimate these expected values, and finally, I discuss how the EKF and the UKF are particular cases of these approximations.

The notation $\mathcal{N}(z;\mu,\Sigma)$ is shorthand for the density of a multivariate normal distribution with argument $z$, mean $\mu$, and covariance $\Sigma$. Let $x_{t|t} \equiv \mathbb{E}[x_t | y_1,...,y_t]$ and $P_{t|t} \equiv Var[x_t | y_1,...,y_t]$. 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 each of the filtering steps are normal. In this case, the conditional density of the state variable $x_t$, denoted by $p(x_t | y_1,...,y_t)$, is characterized by its first and second conditional moments; that is,

$$p(x_t | y_1,...,y_t) \equiv \mathcal{N}\left(x_t; x_{t|t}, P_{t|t}\right).$$

Moreover, the conditional density of $x_{t+1}$ is also Gaussian,

$$p(x_{t+1} | y_1,...,y_t) \equiv \mathcal{N}\left(x_{t+1}; x_{t+1|t}, P_{t+1|t}\right),$$

with conditional moments obtained from the transition equation represented by Eq. (2.2),

$$x_{t+1|t} = \mathbb{E}[g(x_t) | y_1,...,y_t], \tag{2.3}$$

$$P_{t+1|t} = Var[g(x_t) | y_1,...,y_t] + Q. \tag{2.4}$$

Similarly, the measurement density, defined by the observation equation in (2.1) is Gaussian,

$$p(y_{t+1} | y_1,...,y_t) \equiv \mathcal{N}\left(y_{t+1}; y_{t+1|t}, P^{yy}_{t+1|t}\right),$$

with mean

$$y_{t+1|t} = \mathbb{E}[h(x_{t+1}) | y_1,...,y_t] \tag{2.5}$$

and variance–covariance matrix

$$P^{yy}_{t+1|t} = Var[h(x_{t+1}) | y_1,...,y_t] + R, \tag{2.6}$$

where $R$ is the covariance matrix of the measurement shocks. Moreover, the conditional

covariance between the measurements and states is represented by

$$P_{t+1|t}^{xy} = Cov\left[x_{t+1}, h\left(x_{t+1}\right)|y_1, ..., y_t\right]. \tag{2.7}$$

By assuming that the conditional densities are Gaussian, the conditional moments can be obtained recursively by applying the standard Kalman filter,[7] represented by the following set of equations:[8]

$$\begin{aligned} p\left(x_{t+1}|y_1, ..., y_{t+1}\right) &= \mathcal{N}\left(x_{t+1}; x_{t+1|t+1}, P_{t+1|t+1}\right), \\ K_{t+1} &= P_{t+1|t}^{xy}\left(P_{t+1|t}^{yy}\right)^{-1}, \\ x_{t+1|t+1} &= x_{t+1|t} + K_{t+1}\left(y_{t+1} - y_{t+1|t}\right), \\ P_{t+1|t+1} &= P_{t+1|t} - K_{t+1}P_{t+1|t}^{yy}K_{t+1}^{\top}. \end{aligned} \tag{2.8}$$

Equations (2.8) are based on the calculations of the moments of Eqs. (2.3) - (2.7), which are expected values of nonlinear transformations of random variables which may not have a closed form. A natural approach consists of replacing the observation and transition equations with their Taylor series expansions, using its mean vector as the center of expansion of the series. Consequently, the moments of the observation and transition equations involved are calculated with the expected values of the Taylor approximations. In this setup, the EKF corresponds to a first-order approximation, while the UKF coincides with the second-order approximation of the functions that define the state-space model. As a result, the numerical integration problem is solved by calculating the derivatives of the observation and transition equations as well as the cross moments of normally distributed random vectors.

The next section provides the basic setup for estimating the moments of Gaussian random vectors using Taylor approximations. A brief overview of the multivariate version of the Taylor series is presented first, followed by an explanation of how these approximations are used to estimate the moments of possibly nonlinear transformations of normally distributed random vectors.

### 2.3.2 Taylor Series Approximations

Let $y = f\left(\mathbf{x}\right)$ denote a smooth function, where $f : \mathbb{R}^N \longrightarrow \mathbb{R}$, and let $\mu$ denote an $N$-dimensional constant vector.[9] Let $\mathbf{q} = (q_1, ..., q_N)$ denote a vector of nonnegative integers, $|\mathbf{q}| = \sum_{n=1}^{N} q_n$, $\mathbf{q}! = \prod_{n=1}^{N}\left(q_n!\right)$, and $f_{\mathbf{q}}(\mu)$ denote the partial derivative of order $\mathbf{q}$ of the function $f(\mu)$ evaluated at $\mu$; i.e.,

$$f_{\mathbf{q}}\left(\mu\right) = \frac{\partial^{q_1+...+q_N} f}{\partial x_1^{q_1}...\partial x_N^{q_N}}\left(\mu\right). \tag{2.9}$$

[7] A formal proof of this result can be found in Theorem 2 of Kalman (1960).

[8] For a detailed description of the standard Kalman filter, please see Appendix A.3.

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

**Theorem 2.3.1** *Let $U \subset \mathbb{R}^N$ be an open subset, $\mathbf{x} \in U$, $\mu \in \mathbb{R}^N$, so that $t\mathbf{x} + (1-t)\mu \in U$ 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*

$$f(\mathbf{x}) = \sum_{\{\mathbf{q}:|\mathbf{q}|\leq M\}} \frac{1}{\mathbf{q}!} f_{\mathbf{q}}(\mu) \prod_{n=1}^{N} (x_n - \mu_n)^{q_n} + \sum_{|\mathbf{q}|=M+1} \frac{1}{\mathbf{q}!} f_{\mathbf{q}}(\zeta) \prod_{n=1}^{N} (x_n - \mu_n)^{q_n}, \tag{2.10}$$

*where $\zeta = \lambda\mathbf{x} + (1-\lambda)\mu$.*

Theorem 2.3.1 is the preamble to the Taylor series approximations. It shows that $f(\mathbf{x})$ can be rewritten as the sum of a polynomial, where the coefficients are determined by the derivatives of the function evaluated at the point $\mu$, and a term that includes its derivatives of order $M+1$ evaluated at a point $\zeta$. The polynomial is defined as the $M$-th order Taylor approximation to the function $f(\mathbf{x})$, and the second term is known as the remainder.

**Definition** The generic $M$-th order Taylor approximation of $f$ centered at $\mu$, denoted by $\widehat{f}$, is defined as

$$\widehat{f}(\mathbf{x}) = \sum_{\{\mathbf{q}:|\mathbf{q}|\leq M\}} \frac{1}{\mathbf{q}!} f_{\mathbf{q}}(\mu) \prod_{n=1}^{N} (x_n - \mu_n)^{q_n}, \tag{2.11}$$

for $x \in U$.

Now, suppose that $x \sim \mathcal{N}(\mu, \Sigma)$ and that we are interested in calculating the expected value of $f(x)$. A natural approach is to replace the function $f$ with its $M$-th order Taylor approximation and estimate the expected value of this approximation. Thus, from Eq. (2.11) we have

$$\mathbb{E}[f(x)] \simeq \sum_{\{\mathbf{q}:|\mathbf{q}|\leq M\}} \frac{1}{\mathbf{q}!} f_{\mathbf{q}}(\mu)\, \mathbb{E}\left[\prod_{n=1}^{N} (x_n - \mu_n)^{q_n}\right]. \tag{2.12}$$

Intuitively, Eq. (2.12), provides an approximation for the expected value of a transformation of a normally distributed random vector which is based on two separate elements: the derivatives of the function $f$ evaluated at $\mu$ and the cross moments of a normally distributed random vector.[10] In most of the applications, the derivatives of the function $f$ have an analytical expression and can be calculated explicitly. However, the filtering recursions in (2.8) not only involve calculating the expected value of a transformation, they require the calculation of variances and covariances of this transformation with state variables.

The following sections describe how to calculate the cross moments in Eq. (2.12) based on the results of Savits (2006). It is also shown how to apply the Taylor approximations for

[10] Conditions under which $\sum_{|\mathbf{q}|=M+1} \frac{1}{\mathbf{q}!} \mathbb{E}\left[f_{\mathbf{q}}(\zeta) \prod_{n=1}^{N} (x_n - \mu_n)^{q_n}\right] \longrightarrow 0$ as $M \to \infty$, can be found in Jiming (2010) and Garlappi and Skoulakis (2011).

the estimation of variances and covariances. Propositions 2.3.3 and 2.3.4 as well as Lemma 2.3.5 provide the basis for calculating variance–covariance matrices efficiently using Taylor approximations. Finally, all the results are summarized in the nonlinear filter based on Taylor approximations which is presented in Algorithm 2.3.6.

### 2.3.3 Calculation of the Moments of a Multivariate Normal Distribution

Let $Z = (z_1, z_2, ..., z_N)$ denote a multivariate normal random vector with zero-mean vector and covariance matrix $\Sigma$, where the component $i, j$ denotes the covariance between the random variables $z_i$ and $z_j$. Let $\mathcal{M}_{\mathbf{q}}^{\Sigma}$ be the $\mathbf{q} \equiv (q_1, ..., q_N)$ moment, where $q_1, ..., q_N$ are nonnegative integers; i.e., $\mathcal{M}_{\mathbf{q}}^{\Sigma} = \mathbb{E}[z_1^{q_1} ... z_N^{q_N}]$. Then, from Theorem 5.1 in Savits (2006), we have the following recursive relation between the multivariate moments of $Z$.

**Proposition 2.3.2** *Set $\mathcal{M}_{(0,...,0)}^{\Sigma} = 1$; then, for all $\mathbf{q} = (q_1, ..., q_N) \geq \mathbf{0}_N$ and $1 \leq j \leq N$, we have*

$$\mathcal{M}_{\mathbf{q}+\mathbf{e}_j}^{\Sigma} \equiv \mathbb{E}\left[z_1^{q_1} ... z_j^{q_j+1} ... z_N^{z_N}\right] = \sum_{k=1}^{N} \Sigma_{jk} q_k \mathcal{M}_{\mathbf{q}-\mathbf{e}_k}^{\Sigma}, \tag{2.13}$$

*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).*

Proposition 2.3.2 provides a recursive method to estimate the cross moments of a normally distributed random vector based on the variance–covariance matrix only. This result is the basis for the filtering methods via Taylor approximations, because the calculation of moments and derivatives can be separated in a tractable form.

To calculate the second moments involved in the Kalman filter recursions in Eqs. (2.4) and (2.6), I approximate the product of a transformed random variable with its Taylor series around the mean vector $\mu$. The choice of the mean vector, $\mu$, as a center of expansion of the Taylor series is convenient for three reasons: First, all the calculations that involve derivatives are independent of the expectation operator; second, the cross moments are independent of the mean vector; and third, $\mathbb{E}[\prod_{n=1}^{N} (x_n - \mu_n)^{q_n}] = 0$, for all vectors $\mathbf{q}$ such that $\sum_{n=1}^{N} q_n$ is an odd number. In any case, the results will be valid if the center of expansion is any constant vector. If the problem involves the calculation of conditional expectations, then the results will still be valid. The only requirement is for the center of expansion to be measurable with respect to the current information set.

We know that the variance and covariance of any set of random variables involve calculating expectations of the product of random variables. For example, the variance of $f(x)$ requires the

calculation of $\mathbb{E}\left[f^2(x)\right]$ and $\mathbb{E}\left[f(x)\right]$. As described previously, the expectation of $f^2(x)$ can be approximated with its Taylor series; i.e.,

$$\mathbb{E}\left[f^2(x)\right] \approx \sum_{\{\mathbf{q}:|\mathbf{q}|\leq M\}} \frac{1}{\mathbf{q}!}\left(f^2\right)_{\mathbf{q}}(\mu)\,\mathcal{M}_{\mathbf{q}}^{\Sigma}, \tag{2.14}$$

where $\left(f^2\right)_{\mathbf{q}}(\mu)$ denotes the partial derivative of order $\mathbf{q}$ of the function $f^2$ evaluated at $\mu$. Finally, the variance is obtained as the difference between the estimate of the second moment of the function and the squared value of the estimate of the first moment; that is,

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

The same method can be applied to estimate the covariance of two transformed random vectors. The idea is to approximate the expected value of the product of two functions with their derivatives evaluated in the vector of means and the cross moments of a normally distributed random vector. Following Eq. (2.12), we have

$$\mathbb{E}\left[f_1(x)\cdot f_2(x)\right] \approx \sum_{\{\mathbf{q}:|\mathbf{q}|\leq M\}} \frac{1}{\mathbf{q}!}\left(f_1\cdot f_2\right)_{\mathbf{q}}(\mu)\,\mathcal{M}_{\mathbf{q}}^{\Sigma}, \tag{2.15}$$

and the covariances involved in the calculation of the covariance matrix of the observation vector, can be obtained as

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

Clearly, from Eqs. (2.14) and (2.15), we learn that variances and covariances could be estimated with the derivatives of the square and the product of functions which may result in cumbersome calculations. However, Propositions 2.3.3 and 2.3.4 provide a tractable recursive scheme to compute the derivatives of these functions based on the derivatives of the functions $f$, $f_1$ and $f_2$. This is obtained via the Faà di Bruno formula for the derivative of a composite function and its extensions to the multivariate case recently proposed in Savits (2006) and applied in Garlappi and Skoulakis (2010, 2011).[11]

**Proposition 2.3.3** *Let $f:\mathbb{R}^N \longrightarrow \mathbb{R}$ be an $(M+1)$- times continuously differentiable function. Then the derivatives of $\phi(\mathbf{x}) = f(\mathbf{x})^2$ can be obtained from the following vector recursion:*

$$\begin{aligned} \phi_{\mathbf{0}}(\mathbf{x}) &= f(\mathbf{x})^2 \\ \phi_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x}) &= \sum_{\left\{\ell\in\mathbb{N}_0^N:\mathbf{0}_N\leq\ell\leq\mathbf{q}\right\}} 2\times\binom{\mathbf{q}}{\ell} f_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\, f_{\ell}(\mathbf{x}). \end{aligned} \tag{2.16}$$

***Proof*** *See Appendix A.2*

[11] A brief overview of the fundamentals of these recursions and the main results of Savits (2006) are presented in Appendix A.1.

**Proposition 2.3.4** *Let $f_1, f_2 : \mathbb{R}^N \longrightarrow \mathbb{R}$ be $(M+1)$- times continuously differentiable functions. Let $\gamma(\mathbf{x}) = f_1(\mathbf{x}) \cdot f_2(\mathbf{x})$, then the derivatives of $\gamma(\mathbf{x})$, are given by*

$$\begin{aligned} \gamma_{\mathbf{0}}(\mathbf{x}) &= f_1(\mathbf{x}) f_2(\mathbf{x}) \\ \gamma_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x}) &= \sum_{\left\{\ell \in \mathbb{N}_0^N :\, \mathbf{0}_N \leq \ell \leq \mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} f_{1,\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x}) f_{2,\ell}(\mathbf{x}), \\ &\quad + \sum_{\left\{\ell \in \mathbb{N}_0^N :\, \mathbf{0}_N \leq \ell \leq \mathbf{q}\right\}} \binom{\mathbf{q}}{\ell} f_{2,\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x}) f_{1,\ell}(\mathbf{x}). \end{aligned} \tag{2.17}$$

***Proof*** *See Appendix A.2*

Although the use of Taylor series for the calculation of covariances is quite convenient, if either $f_1(\mathbf{x})$ or $f_2(\mathbf{x})$ is linear then we can estimate these covariances in a simpler way. Stein's Lemma provides an expression for the covariance between a normally distributed random vector and its nonlinear transformation. Indeed, Stein's Lemma can be applied to calculate the covariance matrix involved in the Kalman filter step of Equation (2.7).

**Lemma 2.3.5** *(Stein's Lemma). Suppose that $X \equiv (x_1, ..., x_N) \sim \mathcal{N}(\mu, \Sigma)$. 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 $\nabla f(X) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)^\top$. Then the following identity holds:*

$$cov(X, f(X)) = \Sigma \times \mathbb{E}[\nabla f(X)]. \tag{2.18}$$

*Specifically,*

$$cov(x_1, f(x_1, ..., x_N)) = \sum_{i=1}^{N} cov(x_1, x_i) \times \mathbb{E}\left[\frac{\partial}{\partial x_i} f(x_1, ..., x_N)\right]. \tag{2.19}$$

*If $f$ is a vector function, then $\nabla f(X)$ is replaced by the transpose of the Jacobian matrix of $f$, $J_f$ defined by*

$$J_f = \frac{\partial(f_1, f_2, ..., f_p)}{\partial(x_1, x_2, ..., x_n)} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_p}{\partial x_n} \end{bmatrix}.$$

***Proof*** *See Appendix A.2.*

Stein's Lemma involves calculating the expectation of a vector of partial derivatives. If the function $f$ is a polynomial of order less or equal to three, then this expectation is a linear function of the mean vector $\mu$ and the variance–covariance matrix, $\Sigma$. If the function $f$ is not a

polynomial of order less than or equal to three, it is necessary to obtain the Taylor series of each of the components of the vector of partial derivatives. Fortunately, the Taylor approximations of these components are obtained directly from the Taylor series of the function $f$; that is,

$$\frac{\partial f}{\partial x_i} \simeq \sum_{\{\mathbf{q}:0<|\mathbf{q}|\leq M\}} \frac{q_i}{\mathbf{q}!} f_{\mathbf{q}}(\mu) \times (x_i - \mu_i)^{q_i - 1} \times \prod_{\substack{n=1 \\ n \neq i}}^{N} (x_n - \mu_n)^{q_n}. \tag{2.20}$$

Now, taking the expectation of Eq. (2.20) yields to

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

As in Eq. (2.12), calculating the covariance in Eq. (2.18) can be separated in two steps: first the calculation of the derivatives of the vector $\frac{\partial f}{\partial x}$; and, second, estimating the cross moments of a normally distributed random vector. Following the notation of Proposition 2.3.2, we have that the moments required to calculate this covariance are written as

$$\mathcal{M}^{\Sigma}_{\mathbf{q}-\mathbf{e}_i} = \mathbb{E}\left[(x_i - \mu_i)^{q_i - 1} \times \prod_{\substack{n=1 \\ n \neq i}}^{N} (x_n - \mu_n)^{q_n}\right].$$

Therefore, the expectation of each of the partial derivatives is written as

$$\mathbb{E}\left[\frac{\partial f}{\partial x_i}\right] \approx \sum_{\{\mathbf{q}:0<|\mathbf{q}|\leq M\}} \frac{q_i}{\mathbf{q}!} f_{\mathbf{q}}(\mu)\, \mathcal{M}^{\Sigma}_{\mathbf{q}-\mathbf{e}_i}. \tag{2.21}$$

Algorithm 2.3.6 summarizes the filtering method with Taylor series approximations. The main inputs to perform the filtering recursions are the derivatives of the functions that define the observation and transition equations of a state-space model and the cross moments of a multivariate normal distribution. Moreover, the algorithm coincides with the EKF and the UKF when the first-and second-order approximations are considered, respectively.

**Algorithm 2.3.6** *The process is initialized with the unconditional moments of $x_t$:*

$$\begin{aligned} x_{0|0} &= \mathbb{E}[x_0], \\ P_{0|0} &= Var[x_0]. \end{aligned}$$

*For* $t = 0, ..., T$

- *Calculate the derivatives of orders less than or equal to* $M$ *of each of the components of the vector function* $g(x)$, $g_i(x)$, $i = 1, ..., n$, *and evaluate them at the point* $x_{t|t}$.
- *Based on Propositions 2.3.3 and 2.3.4, calculate the derivatives of orders less than or equal to* $M$ *of functions of the form* $g_i(x) g_j(x)$, $i, j = 1, ...n$, *and evaluate them at the point* $x_{t|t}$.
- *Based on Proposition 2.3.2, calculate all the cross moments,* $\mathcal{M}_{\mathbf{q}}^{P_{t|t}}$, *with* $|\mathbf{q}| \leq M$ *of a normally distributed random vector with a mean vector of zero and variance–covariance matrix* $P_{t|t}$.
- *Calculate* $\mathbb{E}[g(x_t)|y_1, ..., y_t]$ *and* $Var[g(x_t)|y_1, ..., y_t]$ *according to Eqs. (2.14) and (2.15).*

***Time Update***

- *Estimate* $x_{t+1|t}$ *and* $P_{t+1|t}$ *as*

$$\begin{aligned} x_{t+1|t} &= \mathbb{E}[g(x_t)|y_1, ..., y_t], \\ P_{t+1|t} &= Var[g(x_t)|y_1, ..., y_t] + Q. \end{aligned}$$

- *Calculate the derivatives of orders less than or equal to* $M$ *of each of the components of the vector function* $h(x)$, $h_i(x)$, $i = 1, ..., p$, *and evaluate at the point* $x_{t+1|t}$.
- *Based on Propositions 2.3.3 and 2.3.4, calculate the derivatives of orders less than or equal to* $M$ *of functions of the form* $h_i(x) h_j(x)$, $i, j = 1, ...n$, *and evaluate them at the point* $x_{t+1|t}$.
- *Calculate all the cross moments,* $\mathcal{M}_{\mathbf{q}}^{P_{t+1|t}}$, *with* $|\mathbf{q}| \leq M$ *of a normally distributed random vector with a mean vector of zero and variance–covariance matrix* $P_{t+1|t}$ *according to Proposition 2.3.2.*
- *Calculate* $\mathbb{E}[h(x_{t+1})|y_1, ..., y_t]$, $\mathbb{E}[\nabla h(x_{t+1})|y_1, ..., y_t]$ *and* $Var[h(x_{t+1})|y_1, ..., y_t]$, *according to Eqs. (2.14), (2.15) and (2.21).*

***Measurement Update***

- *Estimate* $y_{t+1|t}$ *and* $P_{t+1|t}^{yy}$ *as*

$$\begin{aligned} y_{t+1|t} &= \mathbb{E}[h(x_{t+1})|y_1, ..., y_t], \\ P_{t+1|t}^{yy} &= Var[h(x_{t+1})|y_1, ..., y_t] + R, \\ P_{t+1|t}^{xy} &= P_{t+1|t}\mathbb{E}[\nabla h(x_{t+1})|y_1, ..., y_t]. \end{aligned}$$

***Kalman Filter Update***

- *Estimate* $x_{t+1|t+1}$ *and* $P_{t+1|t+1}$ *according to the Kalman update:*

$$\begin{aligned} K_{t+1} &= P_{t+1|t}^{xy}\left(P_{t+1|t}^{yy}\right)^{-1}, \\ x_{t+1|t+1} &= x_{t+1|t} + K_{t+1}\left(y_{t+1} - y_{t+1|t}\right), \\ P_{t+1|t+1} &= P_{t+1|t} - K_{t+1}P_{t+1|t}^{yy}K_{t+1}^{\top}. \end{aligned}$$

From the previous sections, we learned that the filter with Taylor approximations can be applied to a general class of models. However, all these models assume that the state variables are continuous. If a subset of the state variables is discrete, such as the unobserved state variables in a discrete Markov switching model, then the filter with Taylor series approximations cannot be applied directly.

Next section discusses how to apply standard Rao-Blackwelisation methods[12] to the filtering techniques with Taylor approximations. These methods allow including a subset of potentially discrete state variables in a general state-space representation. The approach is done via the Rao-Blackwellised particle filter (Doucet, De Freitas, Murphy, and Russell, 2000). As a result, hybrid filters are obtained where one part of the calculations is performed with the filters based on Taylor approximations and the other part relies on Monte Carlo simulation methods.

### 2.3.4 Rao-Blackwellisation for Filtering with Taylor Series Approximations

Rao-Blackwellised particle filtering was introduced by Doucet, De Freitas, Murphy, and Russell (2000); the basic idea is to reduce the number of variables that must be sampled by identifying variables that have an analytical expression for their density function or can be analyzed in a tractable way. A general overview of the Rao-Blackwellised particle filter can be found in Crisan and Rozovskii (2011).

Suppose that we are interested in analyzing the conditional density of a random vector $x$ that can be written as two different vectors, $z$ and $u$; i.e., $x = [z, u]^\top$. We can think of $z$ as the continuous random variable and $u$ a discrete random vector. The conditional density of $x$ can be written as

$$p(x) = p(z, u) = p(z \,|u)\, p(u). \tag{2.22}$$

If $p(z \,|u)$ admits a closed-form expression, then to approximate the probability density function of $p(x)$ we need to approximate only the unknown density, $p(u)$. In this case, we reduce the number of variables that need to be analyzed. In general, Rao-Blackwellised particle filtering allows to jointly analyze discrete and continuous state variables and reduces the number of variables that must be sampled by taking advantage of the analytical structure of the problem.

Suppose that the state vector $x_t$ can be written as $x_t = (z_t, u_t)^\top$, where $z_t \in \mathbb{R}^N$ and $u_t$ is an unobserved Markov process with known transition probability density and has the following state-space representation:

$$y_t = h(x_t, u_t) + v_t, \qquad v_t \sim \mathcal{N}(0, R(u_t)), \tag{2.23}$$

$$x_{t+1} = g(x_t, u_{t+1}) + \varepsilon_{t+1}, \quad \varepsilon_t \sim \mathcal{N}(0, Q(u_t)), \tag{2.24}$$

[12] See Casella and Robert (1996) for a general reference of this topic.

where $R(u_t)$ and $Q(u_t)$ have the appropriate dimensions. In this case, we can think of $u_t$ as a discrete or continuous state variable that determines a switching state-space model. To solve the filtering problem, we need to estimate a conditional density of the form:

$$p(z_1, ...z_t, u_1, ..., u_t | y_1, ..., y_t) = p(u_1, ..., u_t | y_1, ..., y_t) p(z_1, ...z_t | y_1, ..., y_t, u_1, ..., u_t). \quad (2.25)$$

Conditional upon $(u_1, ..., u_t)$, we have a nonlinear state-space model of the form of Eqs. (2.1) and (2.2). Therefore, to obtain filtered estimates of $z_t$, we first condition upon $(u_1, ..., u_t)$, and obtain $\mathbb{E}(z_t | y_1, ..., y_t, u_1, ..., u_t)$ with the filters based on Taylor approximations as presented in the previous sections. It follows that to fully estimate the unobserved state variables, it remains to approximate the conditional density of $u_1, ..., u_t$ with other filtering approximations such as the particle filter.[13] As a result, we obtain hybrid filters where one part of the calculations is performed analytically and the other part uses Monte Carlo methods or another filter.

The state-space model in Eqs. (2.23) and (2.24) is general enough and can be applied to a number of models such as the partially observed Gaussian models and the Markov switching stochastic volatility models. However, all of these models rely on a set of fixed parameter values that need to be estimated. In section 2.4, I present the quasi-maximum likelihood method for parameter inference of state-space models defined in Section 2.2.

## 2.4 Quasi-Maximum Likelihood Parameter Estimation

Although the focus of the previous sections has been in state estimation, filters based on Gaussian densities, as described in Section 2.3.1 can also be applied for parameter inference of state-space models. White (1982) introduced econometric methods for misspecified models, known as quasi-maximum likelihood (QML) methods. The general idea is to replace the true likelihood function with a Gaussian density and obtain parameter estimates as if the true likelihood function is Gaussian. The parameter estimates are known as QML estimates and can be obtained via standard numerical optimization methods. Moreover, White (1982) shows that QML parameter estimates are statistically consistent. In addition, Gallant and White (1988) provide regularity conditions under which robust standard errors exist.

For dynamic models, such as the state-space models defined in Section 2.2, Bollerslev and Wooldridge (1992) show that by replacing the true likelihood function with the likelihood function constructed with means and covariances obtained from the filtering Eqs. (2.8), QML parameter estimates can be obtained.[14]

[13] Appendix A.6 provides a general introduction to Particle Filters.

[14] For recent applications of these methods in nonlinear state-space representations, see Christoffersen, Jacobs, Karoui, and Mimouni (2012), van Binsbergen and Koijen (2011), Campbell, Sunderam, and Viceira (2011) and Calvet, Fisher, and Wu (2013).

### 2.4.1 Quasi-Likelihood Function

Let $\mathcal{L}(\theta)$ denote the quasi log-likelihood function of a dynamic model evaluated at the vector of parameter values $\theta$; then the function $\mathcal{L}$ is constructed as follows:

For each $t = 1, ..., T$, the conditional mean, $y_{t+1|t}$, and conditional covariance, $P^{yy}_{t+1|t}$, are calculated recursively through Eqs. (2.8). The quasi log-likelihood function is calculated by assuming that $y_{t+1}$, is normally distributed with mean $y_{t+1|t}$ and covariance matrix $P^{yy}_{t+1|t}$; that is,

$$l_t(\theta) = -\frac{p}{2}\ln(2\pi) - \frac{1}{2}\ln\left|P^{yy}_{t+1|t}\right| - \frac{1}{2}\left(y_{t+1} - y_{t+1|t}\right)^{\top}\left(P^{yy}_{t+1|t}\right)^{-1}\left(y_{t+1} - y_{t+1|t}\right). \quad (2.26)$$

Finally, the QML parameter estimates, denoted by $\theta^{QML}$, are obtained by choosing the vector of parameters $\theta$ that maximizes the quasi-likelihood function; that is,

$$\theta^{QML} \equiv \arg\max_{\theta} \mathcal{L}(\theta),$$

where

$$\mathcal{L}(\theta) = \Sigma_{t=1}^{T} l_t(\theta). \quad (2.27)$$

Bollerslev and Wooldridge (1992) show that the QML function, (2.27), is well defined. Moreover, they show that the true but unknown vector of parameters is the global maximizer of (2.27) if the following conditions hold:

$$\begin{aligned} \mathbb{E}[v_{t+1}|y_1, ..., y_t] &= 0, \\ Var[v_{t+1}|y_1, ...y_t] &= R, \end{aligned}$$

which means that if the first and second moments are well specified, then the global maximizer of the QML function will be the true but unknown vector of parameter values. Finally, asymptotic standard errors can be estimated as in Gallant and White (1988). They show that under certain regularity conditions, the covariance matrix of the QML estimator has a closed-form expression as shown in Appendix A.7.

## 2.5 Applications

In this section, I test the filter with Taylor series in three different nonlinear models. The first model is the standard stochastic volatility model as in Andersen and Sørensen (1996). The second model is the risk-return representation analyzed by Brandt and Kang (2004) in the predictability literature, and the third model is a simple version of the dynamic stochastic general equilibrium model studied by Schmitt-Grohe and Uribe (2004). I implement the filter with Taylor series for different approximation orders in simulated and real data. I conduct two sets of exercises. The first set consists of pure filtering exercises that test the accuracy of the

Taylor series for state estimation by using different approximation orders. The second set of exercises analyze the precision of the parameter estimates by comparing the true values with their sample counterparts. Finally, I compare the results with those obtained from standard methodologies such as the EKF, the UKF and the particle filter.

### 2.5.1 Stochastic Volatility Models

The standard stationary stochastic volatility[15] model in discrete time is represented by

$$\begin{aligned} y_t &= \eta_t \cdot \sigma_t \\ \ln \sigma_t^2 &= d + \phi \ln \sigma_{t-1}^2 + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \end{aligned} \tag{2.28}$$

where $y_t$ is the value of a time series observation at time $t$ and $\sigma_t$ is the corresponding volatility, $d$ is a scale parameter for the volatility process and $\eta_t$ is a white noise process with unit variance that represents the innovations in the level or returns. The disturbance of the volatility equation $\varepsilon_t$, is assumed to be a Gaussian white noise process; in addition, $|\phi|$ is considered as a measure of persistence of shocks to the volatility. The variance of the log-volatility process, $\sigma_\varepsilon^2$, 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) which 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 $\phi$ is close to one and $\sigma_\varepsilon^2$ is close to zero, then the evolution of volatility over time is very smooth; however, in the limit, if $\phi = 1$ and $\sigma_\varepsilon^2 = 0$, the volatility is constant over time, and consequently, the returns are homoscedastic. As noted by Broto and Ruiz (2004), if $\sigma_\varepsilon^2 = 0$ the model cannot be identified.

#### State-Space Representation and Implementation

An alternative representation of Eq. (2.28) is obtained by using the demeaned log-volatility, $s_t \equiv \ln \sigma_t^2 - \ln \overline{\sigma}^2$, and $\eta_t$ as state variables. In this case, the re-parameterized stochastic volatility process is

$$y_t = \overline{\sigma} \exp\left(\frac{s_t}{2}\right) \eta_t \tag{2.29}$$

$$s_t = \phi s_{t-1} + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right). \tag{2.30}$$

In terms of the state-space representation in (2.1) and (2.2), the state variables are given by the vector $x_t = [s_t, \eta_t]^\top$, Eq.(2.29) is the observation equation, where $h(x_t) \equiv h(s_t, \eta_t) =$

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

$\overline{\sigma}\exp(s_t/2)\,\eta_t$, and the random noise, $v_t$, as well as its variance, $R$, are equal to zero.[16] Finally, the law of motion is given by two equations. Eq. (2.30) represents the first equation, while the second consists of only one element, a random shock. Overall, the transition equations are represented by

$$\begin{aligned}\begin{bmatrix} s_t \\ \eta_t \end{bmatrix} &= \begin{bmatrix} \phi & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} s_{t-1} \\ \eta_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \quad \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim \mathcal{N}(0, Q) \\ Q &= \begin{bmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & 1 \end{bmatrix},\end{aligned} \tag{2.31}$$

where $\rho \equiv corr(s_t, \eta_t)$. This representation allows capturing the correlation coefficient in a tractable way; since standard state-space representations assume that the shocks between the observation and transition equations are uncorrelated. Moreover, all the moments involved in the filtering recursions exist and have a closed-form expression which allows testing the accuracy of the Taylor series for both filtering and parameter estimation.

Lastly, from Eq. (2.31) we learn that the function that characterizes the transition equation is linear in the state vector, i.e., $g(x_t) \equiv \Phi \cdot x_t$, where

$$\Phi = \begin{bmatrix} \phi & 0 \\ 0 & 0 \end{bmatrix}.$$

According to Section 2.3.2 and Eqs. (2.5)– (2.7), the moments that need to be approximated by the Taylor series are

$$\begin{aligned} y_{t+1|t} &= \mathbb{E}[y_{t+1}|Y_t] \\ &= \mathbb{E}[\overline{\sigma}\exp(s_{t+1})\,\eta_{t+1}|Y_t] \end{aligned}$$

and

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

Finally, for the Kalman update, we need to estimate the covariance matrix between the observation and transition equations, given by

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

[16] This is a unique feature of the Gaussian filters; most of the Monte Carlo filters, such as the particle filter, require that all the variance–covariances involved in the transition and measurement equations should be positive definite.

where the last equality comes from applying Lemma 2.3.5.[17]

### Improved State Identification

A conventional approach to improve state and parameter identification in the stochastic volatility literature[18] is to model the squared value of the observation equation in addition to the observation equation (2.29).[19] As a result, the stochastic volatility model can be represented with the following vector of observables:

$$y_t = \begin{bmatrix} \overline{\sigma} \exp\left(\frac{s_t}{2}\right) \eta_t \\ \overline{\sigma}^2 \exp\left(s_t\right) \eta_t^2 \end{bmatrix}, \tag{2.32}$$

and Eq. (2.31) as the transition equation. The nonlinear function is given by $h\left(x_t\right) \equiv h\left(s_t, \eta_t\right) = \left[\overline{\sigma} \exp\left(s_t/2\right) \eta_t, \overline{\sigma}^2 \exp\left(s_t\right) \eta_t^2\right]^\top$. A standard approach consists of log-linearizing the second component of Eq. (2.32)[20] and performing the state and parameter inference on that version of the model. The main advantage is that the standard model becomes *linear* and the standard Kalman filter can be applied. However, the major disadvantage of log-linearizing the squared equation is that the correlation coefficient, $\rho$, cannot be identified.[21] The advantage of including the squared value of the original observation equation as an observable in Eq. (2.32) is that the information provided by the dynamics of the squared observation can be incorporated directly in the estimation process. Moreover, there is no need log-linearizations, and any information about the correlation coefficient can be incorporated directly in the estimation process.

The first and second moments of Eq. (2.32), can be calculated using Propositions 2.3.3 and 2.3.4 as well as covariance obtained from Stein's Lemma. Finally, I test the accuracy of the Taylor series approximations, by comparing the filtered series of different approximation orders with the one obtained with the closed-form expressions of the moments. [22]

### Monte Carlo Simulation Results

In this section I conduct a Monte Carlo study to test the accuracy of the filters with Taylor approximations for both filtering and parameter inference. For the pure filtering exercise, I first simulate a time series of 500 observations of returns $(y_t)$ and log-volatilities $s_t$ according to Eqs. (2.29) and (2.31), assuming parameter values $\phi = 0.98$, $\sigma_\varepsilon = 0.1414$, $\overline{\sigma} = 1$ and $\rho = -0.5$. These parameter values have been used in Broto and Ruiz (2004), as well as in empirical applications of daily returns.

[17] LemmaA.8.1 in Appendix A.8 provides closed-form expressions for the expected values, variances and covariances involved in the Kalman filter recursions of the stochastic volatility model.

[18] See Koopman and Sandmann (1998) and Broto and Ruiz (2004) for details.

[19] I used the squared of the observation equation as a new observable. However, the cubed or the fourth power could be included as well.

[20] See Ruiz (1994) for details.

[21] Harvey and Shephard (1996) show that the information about the correlation can be recovered from the signs of the innovations.

[22] See Appendix A.8 for details

For each simulated series, I calculate the filtered values for the level $y_t$, as well as for the log-volatility $s_t$, with the filtering techniques based on Taylor approximations with up to a twelfth approximation order using the true parameter values. I denote by TKF-$M$ the filter constructed with $M-th$ order Taylor series approximation, $M = 2, 3, ..., 12$. For comparison purposes, I also filter the same simulated series with the EKF, the UKF,[23] the particle filter with 1000 particles and the filter that uses the closed–form expressions of the moments. I will refer to this filter as the Gaussian filter.[24] Finally, I calculate the mean squared error (MSE) between the simulated series and its filtered counterpart. The experiment is repeated 500 times with a random re-initialization for each run.

Table 2.1 contains means and standard deviations of the MSE of the log-volatility process as well as the MSE of the level of the series. For this model specification, the minimum MSE for both the log-volatility and the observable is obtained with the particle filter, on average, followed by the filters with Taylor approximations of fourth and fifth order. This result is not surprising since the particle filter provides unbiased estimates for both unobserved states and their joint densities. It is worth noting that the statistics of the MSE obtained with the filters of order eight and higher converge to the statistics of the Gaussian filter. Although the UKF is commonly known as a second-order filter, my simulation results show evidence that the MSE results of the second-order filter are slightly smaller than those of the UKF.

As noted from Table 2.1, the particle filter provides the most accurate state estimates. However, to improve the state estimation, I filter the simulated series including the squared value of the first observation equation as a second observation equation and using the true parameter values. The results are shown in Table 2.2. In this case, the filters of order eight or higher provide more accurate state estimates than the particle filter, on average. However, there is not much improvement for the MSE of the observable, $y_t$, which is slightly smaller than the one calculated from the model specification with one equation.

Figures 2.1a–2.1d compare the state estimates of the log-volatility process generated from a single run using different filters. Although the tests are conducted using all the orders of approximation, I only report the eighth order. Figure 2.1a compares the true state with the state estimate and a 95% confidence interval. The true state is represented by a dot, the mean of the filter obtained with Taylor series is given as a dashed line and the solid lines give the 95% confidence interval, constructed as the interval between the 2.5 and 97.5 percentile points. The actual value of the state is within the 95% probability region on approximately 94% of occasions. Figure 2.1b compares the true state with the filter that applies two observation equations. In this case, the actual state is usually very close to the state estimate; on a couple of occasions, the

[23]The UKF parameters were set to $\alpha = 1, \beta = 0$ and $\kappa = 2$; which are optimal for the scalar case.

[24] Ito and Xiong (2000) propose the name of *Gaussian filter* for the filtering methods that use Gaussian densities to approximate the probability density functions involved in the filtering recursions in Eq.(2.8).

actual state is just outside these percentile estimates and the performance is obviously superior to the filter that uses one equation only.

Figure 2.1c compares the actual states with the filtered estimates using the true parameter values. Clearly, the filter that uses two observation equations outperforms the filter that uses only one observation equation. Finally, Figure 2.1d shows the filtered series along with the results of the particle filter. Clearly, both filtered series are very close to the actual state. However, the CPU time used by the particle filter is more than ten times that used to estimate the series with Taylor approximations. To test for the efficiency of the filtering algorithm, I recorded the CPU time needed to estimate the log-volatility filtered series for each order of approximation along with the EKF, the UKF and the particle filter. The results are displayed in Table 2.3. Model 1 represents the model with one observation equation and Model 2, the model with two observation equations. In this case, the filtered series based on Taylor series approximations are calculated at least four times faster than the ones obtained with the particle filter. The filters that use the exact moments, take on average 0.03 and 0.04 seconds to calculate the filtered series, as compared to the 17.87 seconds it takes the particle filter to estimate the same series. The improvement in CPU time as well as the precision of the filtered estimates shows a notable strength of the filters that apply Taylor series with respect to standard particle filters.

As with most nonlinear models it is difficult, if not impossible, to prove that the parameters of a state-space model are uniquely identified. In order to analyze the uniqueness of the QML estimates, I implemented the following procedure. For a fixed vector of parameter values, a path of noisy returns is simulated and a quasi-likelihood function is constructed using the simulated path. An initial identification exercise is performed by calculating the quasi-likelihood function of a fifth order of approximation in the set of parameters used for the simulation and changing independently each parameter $\phi$, $\sigma_\varepsilon$, $\overline{\sigma}$ and $\rho$. These plots are known as quasi-likelihood contours and are shown in Figure 2.2. The dashed lines represent the *unknown* parameter values that are used to simulate the data. If the parameter is well identified, then the quasi-likelihood function should achieve a maximum parameter value that is close enough to the one used to simulate the data. For comparison I show the quasi-likelihood functions of the EKF, the UKF and the Gaussian filter. The concavity of the quasi-likelihood function with respect to each parameter evidence that all the parameters are well identified. For this specific exercise, $\phi$ is well identified by all the filters. However, $\sigma_\varepsilon$, $\overline{\sigma}$ and $\rho$ are only identified properly by the fifth-order and Gaussian filters. For these parameters, the quasi-likelihood function of the EKF and UKF provide biased parameter values, since the maximum value is far from its true value. Moreover, Figure 2.2c shows a constant value for the quasi-likelihood function of $\overline{\sigma}$, showing that the EKF is not able to identify it.

A second alternative to analyze the finite sample properties of the QML estimator is via Monte Carlo simulations. In particular, I estimate the model from 250 independent samples of

$T = 500$ using QML estimation methods. For each simulation, I construct QML functions of orders three, five, seven and nine. I optimize the function numerically and obtain parameter estimates. As a starting point for the parameter estimates, I chose the true parameter value used to simulate the data.[25] I conduct the same exercise with the EKF, the UKF and the Gaussian filter and present the results in Table 2.4. The table shows the true parameter values in the first row as well as the sample mean and standard deviation of the corresponding parameter estimates in parentheses. Clearly, the average estimates are all close to the true parameter values, suggesting that the QML estimates via Taylor series are relatively unbiased while the average estimate of $\overline{\sigma}$ under the EKF, the UKF and the third-order approximations are biased. Moreover, the standard deviations of the QML estimates of the ninth-order approximation are high and coincide with the standard errors of the third-order approximation. This may be caused by the small sample bias as well as the effect of numerical errors.

**Consumption Growth Model: Estimation**

From the Monte Carlo simulation exercises, we learned that the filter with Taylor approximations is a good approach for parameter estimation of the standard stochastic volatility model. However, the results may differ in historical data in which there is substantial evidence of stochastic volatility. As a robustness check, I perform a parameter estimation exercise with real consumption growth data and estimate a stochastic volatility model with different approximation orders.

A number of models, including Tallarini (2000) and Barillas, Hansen, and Sargent (2009), assume that the log-consumption growth, denoted by $\Delta \ln(C_{t+1})$, follows a random walk with drift $\mu_c$ and standard deviation, $\overline{\sigma}$,

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

However, there is substantial evidence of time variation in the conditional standard deviation of many macroeconomic series (Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry, 2012; Clark, 2009; Fernández-Villaverde and Rubio-Ramírez, 2007; Justiniano and Primiceri, 2008; McConnell and Perez-Quiros, 2000; Stock and Watson, 2002).

As a result, Bidder and Smith (2011) propose an alternate endowment process that features stochastic volatility in log-consumption growth that is consistent with (2.29). The model is

[25] This point is used merely for convenience. However, as a robustness check, I conducted a similar estimation exercise that uses random vectors as starting points for the numerical optimization procedure. In all the cases, the vector of parameter values is fully identified and the average estimates are similar to the ones obtained with the true parameter value as initial point.

represented as follows:

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

where $s_t$ represents the log-volatility processes.

I analyze the performance of the filtering technique by implementing the stochastic volatility model in (2.29) and (2.30) with consumption growth data. I use the monthly series from the Federal Reserve Bank of Philadelphia to construct the real consumption per capita from January 1959 to March 2012. I construct the *monthly* log-consumption growth data 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. This real-time data set of macroeconomic 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.[26] Summary statistics of the time series of monthly consumption growth are shown in Table 2.5.

Quasi-likelihood parameter estimation is performed to the data with the filters of orders three, five, seven and nine, along with the estimates obtained with the EKF, the UKF and the Gaussian filters. The choice of the starting point used for the numerical optimization is as follows. First, I simulated 100 random vectors from the parameter space. Each simulated vector was used as an initial point for the numerical maximization of the quasi-likelihood function constructed with the monthly consumption growth data. Finally, the initial point used to estimate the parameter values is the average of the parameter estimates obtained from the previous procedure.

The results are shown in Table 2.6. In this case, all the parameters are identified. The parameter estimates using the quasi-likelihood function constructed with an order of three or higher have similar values; however, the adjusted standard errors change as the order of approximation changes, mainly caused by numerical errors of the second derivative of the quasi-likelihood function. In general, the magnitude of the standard errors decreases as the order of approximation increases. Overall, the parameter estimates are consistent with 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, which include the recession starting in the last quarter of 2007.

### 2.5.2 Risk and Return Model

Brandt and Kang (2004) introduce a nonlinear representation for the return dynamics that

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

allows for positive risk premium in the context of a latent vector autoregressive (VAR) process. Let $y_t$ be the continuously compounded excess returns with time-series dynamics represented by

$$y_t = \mu_{t-1} + \sigma_{t-1}\varepsilon_t \text{ with } \varepsilon_t \sim \mathcal{N}(0,1), \tag{2.33}$$

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

$$\begin{bmatrix} \ln \mu_t \\ \ln \sigma_t \end{bmatrix} = d + A \begin{bmatrix} \ln \mu_{t-1} \\ \ln \sigma_{t-1} \end{bmatrix} + \eta_t \text{ with } \eta_t \sim \mathcal{N}(0, \Sigma), \tag{2.34}$$

where

$$\begin{aligned} d &= \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}, \; A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ and} \\ \Sigma &= \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \text{ with } b_{12} = b_{21} = \rho\sqrt{b_{11}b_{22}}. \end{aligned} \tag{2.35}$$

The first equation of the VAR in (2.34) describes the dynamics of the logarithm of the conditional mean, and it captures the permanent and temporary components as shown in Fama and French (1988b) and Lamoureux and Zhou (1996), 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 includes the standard stochastic volatility model. Setting $a_{21} = 0$ is the stochastic volatility model estimated by Andersen and Sørensen (1996), Kim, Shephard, and Chib (1998), Jacquier, Polson, and Rossi (2004) and Jacquier, Johannes, and Polson (2007). The latent VAR approach in Eqs. (2.34)−(2.35) allows us to study the contemporaneous and intertemporal relationships between expected returns and risk without relying on predictors. The contemporaneous relationship between the conditional mean and volatility is captured by the correlation coefficient $\rho$, while the intertemporal relationships between expected returns and volatilities are captured by the coefficient matrix $A$.

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

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

and

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

where $\otimes$ represents the Kronecker product.

The return dynamics presented in Eq. (2.34) has two key elements: the transition matrix

$A$ and the correlation coefficient $\rho$. 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 can be incorporated by allowing the conditional mean and volatility to be correlated with the return innovations, denoted by $Corr\left[\varepsilon_t, \eta_t\right] = \left[\rho_\mu, \rho_\sigma\right]^\top$. Different models can be specified by setting these correlations to zero. The assumptions about the correlation between the return innovations and the conditional moment innovations serve to capture and potentially distinguish two popular explanations of asymmetric volatility. Asymmetric volatility refers to the empirical finding that increases in volatility tend to be associated more often with large negative returns than with equally large positive returns. The two popular explanations of asymmetric volatility are the leverage effect and the volatility feedback effect. The leverage effect states that when the value of a firm drops given a large negative return, the leverage of the firm and the associated probability of bankruptcy increase, causing the equity claims to become riskier. The volatility feedback effect attributes the asymmetric volatility to the equilibrium response of the conditional mean to changes in volatility.

**State-Space Representation and Implementation**

The representation given in Eqs. (2.33) and (2.34) defines a state-space model, in which the first equation is the nonlinear measurement equation and the second equation is a linear transition equation. To make inferences about expected returns, volatilities and the parameters of the VAR based on the observed returns, a nonlinear filtering problem needs to be solved. The solution to the nonlinear filtering problem generates estimates of expected log-returns and log-volatilities, $\mathbb{E}\left[\ln \mu_t, \ln \sigma_t \,|y_1, ..., y_t\right]$, as well as variances $Var\left[\ln \mu_t, \ln \sigma_t \,|y_1, ..., y_t\right]$. As in the stochastic volatility model, parameter inference can be performed with QML methods.

A simpler representation of the state-space model can be obtained by redefining the state variables in demeaned terms; that is, $m_t = \ln \mu_t - \ln \overline{\mu}$ and $v_t = \ln \sigma_t - \ln \overline{\sigma}$, 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, by rewriting Eqs. (2.33) and (2.34) in terms of new state variables, the state-space becomes standard in the sense that all the state variables are observed at time $t$. Let $x_t = \left[x_{1t}, x_{2t}, x_{3t}, x_{4t}, x_{5t}\right]^\top = \left[m_{t-1}, v_{t-1}, \varepsilon_t, m_t, v_t\right]^\top$; then equations (2.33) and (2.34) can be rewritten as:

$$y_t = \overline{\mu} \exp\left(x_{1t}\right) + \overline{\sigma} \exp\left(x_{2t}\right) x_{3t}, \tag{2.36}$$

and

$$x_t = \widetilde{A} x_{t-1} + \Gamma \underline{w_t} \text{ with } \underline{w_t} \sim \mathcal{N}\left(0, \Sigma\right), \tag{2.37}$$

where

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

and

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

where $Corr\left[\varepsilon_t, \eta_t\right] = \left[\rho_\mu, \rho_\sigma\right]^\top$. Clearly, $h\left(x_t\right) = \overline{\mu} \exp\left(x_{1t}\right) + \overline{\sigma} \exp\left(x_{2t}\right) x_{3t}$ and, as in the stochastic volatility model, the random noise of the observation equation, $v_t$, is identically zero. As a result the variance of the observation equation is zero $\left(R \equiv 0\right)$.[27] The transition equation is linear; i.e., $g\left(x_t\right) = \widetilde{A} x_t$.

The filtering and parameter estimation problem can be solved using the results of Section 2.3.2. In this case, the first and second moments that need to be calculated are

$$y_{t+1|t} = \mathbb{E}\left[y_{t+1} \left| Y_t \right.\right] = \overline{\mu} \mathbb{E}\left[\exp\left(x_{1t+1}\right) \left| Y_t \right.\right] + \overline{\sigma} \mathbb{E}\left[\exp\left(x_{2t+1}\right) x_{3t+1} \left| Y_t \right.\right], \tag{2.38}$$

and

$$P^{yy}_{t+1|t} = Var\left[y_{t+1} \left| Y_t \right.\right] = \mathbb{E}\left[y^2_{t+1} \left| Y_t \right.\right] - y^2_{t+1|t}. \tag{2.39}$$

The covariance term involved in the Kalman gain is calculated as

$$P^{xy}_{t+1|t} = cov \begin{bmatrix} x_{1t}, y_{t+1} \left| Y_t \right. \\ x_{2t}, y_{t+1} \left| Y_t \right. \\ x_{3t}, y_{t+1} \left| Y_t \right. \end{bmatrix} = P_{t+1|t} \times \mathbb{E} \begin{bmatrix} \overline{\mu} \exp\left(x_{1t}\right) \left| Y_t \right. \\ \overline{\sigma} \exp\left(x_{2t}\right) x_{3t} \left| Y_t \right. \\ \overline{\sigma} \exp\left(x_{2t}\right) \left| Y_t \right. \end{bmatrix}, \tag{2.40}$$

by applying Lemma 2.3.5. Closed-form expressions for the moments of Eqs. (2.38) - (2.40) can be obtained from Proposition A.8.1 in Appendix A.8.

**Improved State Identification**

As in the stochastic volatility example and following Brandt and Kang (2004), state identification can be improved by including additional observation equations to the filtering problem. An

[27] The main purpose of this representation is to allow for a general correlation structure between the shocks of the observation equation and transition equations, which in standard state-space representations are assumed to be zero.

alternative is to include the squared value of Eq. (2.36), and the vector of observables becomes

$$y_t = \begin{bmatrix} \overline{\mu} \exp(x_{1t}) + \overline{\sigma} \exp(x_{2t})\, x_{3t} \\ (\overline{\mu} \exp(x_{1t}) + \overline{\sigma} \exp(x_{2t})\, x_{3t})^2 \end{bmatrix} \tag{2.41}$$

with Eq. (2.37) as the transition equation of the state-space representation. Other equations that can be included are the third observation equation, given by the cube of Eq.(2.36) or the product of the current observed return with its lagged value. By including an extra observation equation in Eq. (2.41), we need to estimate its mean and all its covariances to have estimates of $y_{t+1|t}$, $P^{yy}_{t+1|t}$ and $P^{xy}_{t+1|t}$, which can be done via the Taylor series.

**Monte Carlo Simulation Results**

Accuracy in state estimation is one of the most important features of any filtering technique. To test the accuracy of the filters with Taylor approximations in the model represented in Eqs. (2.36)-(2.37), I use Monte Carlo simulations. First, using the parameter estimates of Brandt and Kang (2004) for Model A, I simulate a random path of returns and calculate the filtered series of both log-expected returns and log-volatilities using different orders of approximation and the true parameter values. The pure filtered series of log-expected returns and log-volatilities are almost constant, showing an average correlation of 0.05%. To improve the state estimation, I filter the same series using the vector of observation equations (2.41). The results of a random simulation are shown in Figure 2.3. The filtered log-returns are shown in the top left (Figure 2.3a) while the filtered log-volatilities are displayed in the top right (Figure 2.3b). In both cases, the filtered series based on one observation equation provide almost constant estimates. However, the filtered series based on two observation equations provide a much more accurate estimate. The average correlation coefficient between the true states and the filtered series is above 90%. As a robustness check, I include the filtered series from a model that has the cubed observation equation. As shown in Figures 2.3a and 2.3b, the difference between the model with two observation equations and the model with three observation equations is indistinguishable.

To test the accuracy of the filter with respect to the particle filter, based on a simulated series, I estimate the filtered series of the model with two observation equations using the true parameter values and contrast the estimates with the ones obtained with the particle filter. As shown in Figures 2.3a and 2.3b, both methods provide similar values for the filtered log-expected returns and log-volatilities.

To get a sense of the accuracy of the Taylor approximations, I simulated a random path of $T = 792$ using the parameter estimates by Brandt and Kang (2004). Using this simulation, I calculated the value of the quasi-log likelihood function of the model in Eqs. (2.36) and (2.37) in the true vector of parameter values for up to fifteen orders of approximation. For comparison purposes, I include the quasi-likelihood function obtained with the UKF and the one constructed

with the exact moments. Figure 2.4 shows 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. In this case, a Taylor series approximation of order six, is necessary to get an estimate of the likelihood function close enough to the one estimated with the exact moments.

As with the stochastic volatility model, it is a challenge to prove that the parameters of the model are well identified. As an identification exercise and assuming $\rho_\mu = \rho_\sigma = 0$, I simulate a sample path of stock returns with $T = 5000$ using the parameter estimates from Brandt and Kang (2004). The quasi-likelihood function is constructed with a degree of approximation of $M = 6$. The function is evaluated numerically at the true parameter values, by varying one parameter at a time and keeping the remaining fixed. The plot of the parameter value versus the likelihood function is known as the likelihood contour. If the parameter is well identified, then the plot will have a concave shape and the maximum is achieved at the *true* parameter value. If the shape of this function is constant or a straight line, then there is evidence of misidentification. The results are shown in Figure 2.5; the dashed lines represent the *unknown* parameter values used to generate the data. The concavity of the quasi-likelihood contours is evidence that all the parameters are well identified and the maxima are achieved at values close enough to the ones used to simulate the data. It is worth mentioning that the number of observations, $T$, is important. By using a smaller number of observations, the correlation coefficients may not be identified.

To provide more evidence of the accuracy of QML estimates, I study four different assumptions about the correlation structure between return innovations and the conditional mean and volatility innovations. Model A assumes that these innovations are uncorrelated; in Models B and C, the return innovations are allowed to be correlated either with the conditional mean or the conditional volatility innovations, respectively; and lastly, Model D is unrestricted. For each model, I simulate 500 independent samples of $T = 792$ monthly returns using the parameter values obtained of Brandt and Kang (2004) and I obtain the QML parameter estimates for different orders of approximation.[28] I report the results only for $M = 6$, since Figure (2.4) shows that an order of $M = 6$ is necessary to obtain an accurate likelihood.

Table 2.7 presents the results for the four models. For each model, I show the parameter values used to simulate the data as well as the sample mean and standard deviation of the QML estimates. Overall, the parameter estimates show evidence of consistency. However, the small number of simulations does not provide an accurate assessment of the asymptotic unbiasedness. In general, 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. The correlation coefficients, $\rho$ and $\rho_\mu$,

[28] As in the stochastic volatility model, the starting point for the estimation exercise was the true vector of parameter values. As a robustness check, I estimated numerically the models using random vectors as starting points and obtained similar results.

in general, are not identified.[29] A common approach to correct this issue is to add another observation equation, such as the squared value of returns, or to include a set of predictors in the dynamics of the mean and volatility of returns (Brandt and Kang, 2004; Lundblad, 2007).

**Market Excess Returns: Estimation**

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 2.8 presents summary statistics of the data and Figure 2.7 plots the time series of the market portfolio (top) and the short rate (bottom).

**Parameter Estimates**

The model in Eqs. (2.36)-(2.37) is estimated with QML methods using as starting point of the numerical optimization the parameter estimates of Brandt and Kang (2004). Table 2.9 displays the results of the four models. Clearly, the contemporaneous correlation between the conditional mean and volatility, $\rho$, is negative and statistically significant for all the models. As a result, I strongly reject the hypothesis of the lack of contemporaneous relationship between the conditional mean and the conditional volatility. In addition, the correlations between the return innovations and the mean and volatility innovations ($\rho_\mu$ and $\rho_\sigma$) are negative and significant. These results are consistent with French, Schwert, and Stambaugh (1987), Campbell and Hentschel (1992a) and Brandt and Kang (2004). The parameter estimates for the transition matrix $(a_{11}, a_{12}, a_{21}, a_{22})$, are robust under the four specifications. However, the estimates of the standard deviation of the conditional mean and volatility ($b_{11}$ and $b_{22}$) differ between the specifications in models A and B and the ones of models C and D.

### 2.5.3 A Dynamic Stochastic General Equilibrium Model

In this section I show how filtering with Taylor series facilitates quasi-likelihood-based inference in dynamic equilibrium models. I describe how to use the filter with Taylor series to estimate the structural parameters of the model, those characterizing preferences and technology, based on macroeconomic variables measured with noise. Flury and Shephard (2011) suggest that particle filters are the only feasible approach to estimating parameters of dynamic stochastic general equilibrium (DSGE) models. I show that filtering techniques based on Taylor series are an alternative approach for state and parameter estimation in a DSGE model without relying on Monte Carlo filters. Finally, I illustrate the technique with a very simple real business cycle model.

Likelihood-based inference is a useful tool to take dynamic equilibrium models to the data (An and Schorfheide, 2007). However, most dynamic equilibrium models do not

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

imply a likelihood function that can be evaluated analytically or numerically. To circumvent this problem, the literature has used the approximated likelihood derived from a linearized version of the model, instead of the exact likelihood. But linearization depends on the accurate approximation of the solution of the model by a linear relation. This assumption is arguable. First, the impact of linearization is more problematic than it appears. Fernández-Villaverde, Rubio-Ramírez, and Santos (2006) prove that second-order approximation errors in the solution of the model have first-order effects on the likelihood function. Moreover, the error in the approximated likelihood gets compounded with the size of the sample. Period by period, small errors in the policy function accumulate at the same rate at which the sample size grows. Therefore, the likelihood implied by the linearized model diverges from the likelihood implied by the exact model. Fernández-Villaverde and Rubio-Ramírez (2007) document how those insights are quantitatively relevant to real-life applications.

Filters based on Taylor approximations are an alternative approach for filtering and parameter estimation of DSGE models, as they integrate higher order approximations for solutions of DSGE models, such as perturbations, with QML methods. In this section I introduce a simple model that illustrates this method. Based on Monte Carlo simulations, I illustrate the accuracy of the filter in both state and parameter estimation.

**The Model**

In this setup, it is assumed that there is a representative household maximizing its lifetime utility given by

$$\mathbb{E}_0 \left[ \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma} \right], \quad \beta \in (0,1), \quad \gamma > 0, \tag{2.42}$$

where $C_t$ is consumption at time $t$, $\beta$ is the subjective discount factor and $\gamma$ is the risk aversion parameter.

In this economy, there is a production sector, where the date $t$ output flow $Y_t$, is related to date $t$ level of the capital stock, $K_t$, via

$$Y_t = A_t K_t^{\alpha}, \tag{2.43}$$

where $A_t$ is an exogenous technological shock, given by a first-order autoregressive process; that is,

$$\ln A_{t+1} = \rho \ln A_t + \varepsilon_{t+1}, \quad \varepsilon_t \sim N\left(0, \sigma_A^2\right). \tag{2.44}$$

The stock of capital in the date $t$ is related to the date $t$ investment flow $I_t$ and the capital depreciation rate $\delta$ via the standard capital accumulation equation,

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

The aggregate resource constraint is

$$C_t = I_t + Y_t.$$

The central planner's problem consists of choosing $C_t$ and $K_t$ that maximize the expected utility of the form

$$\max_{\{C_t, K_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \left[ \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma} \right], \quad \beta \in (0,1), \quad \gamma > 0 \tag{2.45}$$

subject to

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

and (2.43), for $t = 0, 1, ...$; $K_0, A_0$ given.

**Characterizing the Solution**

The first-order condition implied by (2.43), (2.44) and (2.45) yields to the following Euler equation,

$$C_t^{-\gamma} = \beta \mathbb{E}_t \left[ C_{t+1}^{-\gamma} \left( 1 - \delta + \alpha A_{t+1} K_{t+1}^{\alpha-1} \right) \right], \tag{2.46}$$

while

$$K_{t+1} = A_t K_t^{\alpha} + (1-\delta) K_t - C_t, \tag{2.47}$$

$$\ln A_{t+1} = \rho \ln A_t + \varepsilon_{t+1}, \tag{2.48}$$

are the constraints implied by the model. Equations $(2.46) - (2.48)$ fully characterize the solution of the optimization problem faced by the central planner.

The solution to the system in (2.46) consists of finding policy functions $\pi$ and $\phi$ such that

$$\begin{aligned} C_t &= \pi(K_t, A_t, \chi) \\ \begin{bmatrix} K_{t+1} \\ \log A_{t+1} \end{bmatrix} &= \begin{bmatrix} \phi(K_t, A_t, \chi) \\ \log(A_t) \end{bmatrix} + \chi \begin{bmatrix} 0 \\ \sigma_A \end{bmatrix} \varepsilon_{t+1}, \end{aligned}$$

where $\chi$ is a perturbation parameter. As $\chi \to 0$, the dynamic system in $(2.46) - (2.48)$ converges to a point known as the non-stochastic steady state. However, the system of functional equations implied by the equilibrium conditions does not have, in general, an analytic solution. An approximate solution for the policy functions can be obtained via perturbation methods.[30] The main objective of these methods is to estimate values of the derivatives of $\pi$ and $\phi$ at the non-stochastic steady state. For analytical convenience, the system is written in terms of the

[30] See Judd (1998) and DeJong and Dave (2011) for a detailed explanation of perturbation methods in Economics.

log-deviations of the non-stochastic steady state. Let

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

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

$$C_{ss} = (K_{ss})^{\alpha} - \delta K_{ss}, \tag{2.49}$$

$$K_{ss} = \left[\frac{\alpha\beta}{1-\beta\left(1-\delta\right)}\right]^{\frac{1}{1-\alpha}}. \tag{2.50}$$

The general idea of perturbation methods is to provide a Taylor expansion of the policy functions that characterize the equilibrium of the economy in terms of the state variables of the model and a perturbation parameter, $\chi$. In this case, I construct a second-order approximation for $\widehat{c}_t$ and $\widehat{k}_{t+1}$ of the form

$$\begin{aligned}\widehat{k}_{t+1} &= \phi_k\widehat{k}_t + \phi_a\widehat{a}_t + \frac{1}{2}\left(\phi_{kk}\widehat{k}_t^2 + 2\phi_{ak}\widehat{a}_t\widehat{k}_t + \phi_{aa}\widehat{a}_t^2\right) + \frac{1}{2}\phi_{\chi\chi}\chi^2, \qquad (2.51)\\ \widehat{c}_t &= \pi_k\widehat{k}_t + \pi_a\widehat{a}_t + \frac{1}{2}\left(\pi_{kk}\widehat{k}_t^2 + 2\pi_{ak}\widehat{a}_t\widehat{k}_t + \pi_{aa}\widehat{a}_t^2\right) + \frac{1}{2}\pi_{\chi\chi}\chi^2,\end{aligned}$$

and

$$\widehat{a}_{t+1} = \rho\widehat{a}_t + \varepsilon_{t+1}, \tag{2.52}$$

where $\phi_k, \phi_a, \phi_{kk}, \phi_{ak}, \phi_{aa}, \phi_{\chi\chi}, \pi_k, \pi_a, \pi_{kk}, \pi_{ak}, \pi_{aa}$ and $\pi_{\chi\chi}$ denote the first- and second-order derivatives of the functions $\phi$ and $\pi$, respectively.[31] These derivatives are calculated numerically as in Schmitt-Grohe and Uribe (2004).[32]

**State-Space Representation and Implementation**

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

$$\widehat{GDP}_t = \widehat{y}_t + v_{y,t}, \quad v_{y,t} \sim \mathcal{N}\left(0, \sigma_y^2\right), \tag{2.53}$$

where $v_{y,t}$ represents the measurement error. Now, from Eq. (2.43) we know that the log-GDP implied by the model is

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

[31] For simplicity, I apply second-order approximations to solve the model. However, the following results can be generalized to the perturbation of any order.

[32] I thank Stephanie Schmitt-Grohe and Martin Uribe for making their code available.

Additionally, I assume that $\widehat{k}_t$ and $\widehat{a}_t$ are the unobservable state variables and the econometrician wants to make inference of the state variables as well as the parameters based solely on the $\widehat{GDP}_t$ observations available. Consequently, a nonlinear filtering problem has to be solved.

In this case, Eqs. $(2.51) - (2.53)$ define a state-space model of the form:

$$\begin{aligned} \widehat{GDP}_t &= \alpha\widehat{k}_t + \widehat{a}_t + v_{y,t},\ v_{y,t} \sim \mathcal{N}\left(0, \sigma_y^2\right), \\ \widehat{k}_{t+1} &= \phi_k\widehat{k}_t + \phi_a\widehat{a}_t + \frac{1}{2}\left(\phi_{kk}\widehat{k}_t^2 + 2\phi_{ak}\widehat{a}_t\widehat{k}_t + \phi_{aa}\widehat{a}_t^2\right) + \frac{1}{2}\phi_{\chi\chi}\chi^2, \\ \widehat{a}_{t+1} &= \rho\widehat{a}_t + \varepsilon_{t+1},\ \varepsilon_t \sim \mathcal{N}\left(0, \sigma_A^2\right). \end{aligned} \tag{2.54}$$

The state variables of the filtering problem are represented by the vector $x_t = \left[\widehat{k}_t, \widehat{a}_t\right]^\top$. The observation equation is linear; i.e., $h\left(x_t\right) \equiv h\left(\widehat{k}_t, \widehat{a}_t\right) = \alpha\widehat{k} + \widehat{a}_t$ with a random noise, $v_{y,t}$, with variance $R \equiv \sigma_y^2$. The transition equation is characterized by the nonlinear mapping $g\left(x_t\right)$ represented by the vector:

$$g\left(x_t\right) \equiv g\left(\widehat{k}_t, \widehat{a}_t\right) = \begin{bmatrix} \phi_k\widehat{k}_t + \phi_a\widehat{a}_t + \frac{1}{2}\left(\phi_{kk}\widehat{k}_t^2 + 2\phi_{ak}\widehat{a}_t\widehat{k}_t + \phi_{aa}\widehat{a}_t^2\right) + \frac{1}{2}\phi_{\chi\chi}\chi^2 \\ \rho\widehat{a}_t \end{bmatrix}. \tag{2.55}$$

The first component of the vector in Eq. (2.55) is a quadratic function of the state variables, while the second component of $g\left(x_t\right)$ is a linear function of $\widehat{a}_t$ only. Finally, the variance that characterizes the shock of the transition equation is defined by $\sigma_A^2$; i.e., $Q \equiv \sigma_A^2$.

As the observation and transition equations of the model are polynomials, their expected values will coincide with the expected values calculated with the Taylor series expansions, as long as the order of the polynomial is smaller than the order of the Taylor approximation. According to Section 2.3.2 and Eqs. $(2.5) - (2.7)$, the mean vector of state variables,

$$x_{t+1|t} = \begin{bmatrix} \mathbb{E}\left[\widehat{k}_{t+1}\left|Y_t\right.\right] & \mathbb{E}\left[\widehat{a}_{t+1}\left|Y_t\right.\right] \end{bmatrix}^\top \tag{2.56}$$

is computed by applying the second-order Taylor series approximations of Eq. (2.55). Now, since the transition equation is quadratic, its variance requires fourth-order polynomials; as a result, the variance of the transition equation is calculated exactly with a fourth-order Taylor series, as

$$P_{t+1|t} = \begin{bmatrix} Var_t\left[\widehat{k}_{t+1}\left|Y_t\right.\right] & cov_t\left[\widehat{k}_{t+1}, \widehat{a}_{t+1}\left|Y_t\right.\right] \\ cov_t\left[\widehat{k}_{t+1}, \widehat{a}_{t+1}\left|Y_t\right.\right] & Var_t\left[\widehat{a}_{t+1}\left|Y_t\right.\right] + \sigma_A^2 \end{bmatrix}, \tag{2.57}$$

where each of the components can be calculated using the results from Section 2.3.2. The observation equation is linear in the state variables; therefore, Eqs. $(2.5) - (2.6)$ are given by

$$y_{t+1|t} = \mathbb{E}\left[y_{t+1}\left|Y_t\right.\right] = [\alpha, 1] \cdot x_{t+1|t} \tag{2.58}$$

and

$$P_{t+1|t}^{yy} = [\alpha, 1] \cdot P_{t+1|t} \cdot [\alpha, 1]^\top + \sigma_y^2. \tag{2.59}$$

Now, from Lemma 2.3.5 the covariance between the observation and transition equation is

$$P_{t+1|t}^{xy} = P_{t+1|t} \cdot [\alpha, 1]^\top. \tag{2.60}$$

**Improved State and Parameter Identification**

For inference purposes, a second observation equation can be included. A natural observable to include is consumption per capita, which is measured with *noise*,

$$\widetilde{C_t} = \widehat{c}_t + v_{c,t}, \quad v_{c,t} \sim \mathcal{N}\left(0, \sigma_c^2\right), \tag{2.61}$$

where

$$\widehat{c}_t = \pi_k \widehat{k}_t + \pi_a \widehat{a}_t + \frac{1}{2}\left(\pi_{kk}\widehat{k}_t^2 + 2\pi_{ak}\widehat{a}_t\widehat{k}_t + \pi_{aa}\widehat{a}_t^2\right) + \frac{1}{2}\pi_{\chi\chi}\chi^2.$$

In this case, the state-space model becomes

$$\begin{aligned}
\widehat{GDP_t} &= \alpha\widehat{k}_t + \widehat{a}_t + v_{y,t}, \quad v_{y,t} \sim \mathcal{N}\left(0, \sigma_y^2\right), \\
\widetilde{C_t} &= \pi_k \widehat{k}_t + \pi_a \widehat{a}_t + \frac{1}{2}\left(\pi_{kk}\widehat{k}_t^2 + 2\pi_{ak}\widehat{a}_t\widehat{k}_t + \pi_{aa}\widehat{a}_t^2\right) + \frac{1}{2}\pi_{\chi\chi}\chi^2 + v_{c,t}, \quad v_{c,t} \sim \mathcal{N}\left(0, \sigma_c^2\right), \\
\widehat{k}_{t+1} &= \phi_k \widehat{k}_t + \phi_a \widehat{a}_t + \frac{1}{2}\left(\phi_{kk}\widehat{k}_t^2 + 2\phi_{ak}\widehat{a}_t\widehat{k}_t + \phi_{aa}\widehat{a}_t^2\right) + \frac{1}{2}\phi_{\chi\chi}\chi^2, \\
\widehat{a}_{t+1} &= \rho\widehat{a}_t + \varepsilon_{t+1}, \varepsilon_t \sim \mathcal{N}\left(0, \sigma_A^2\right).
\end{aligned} \tag{2.62}$$

Clearly, the state-space in Eq. (2.62) is nonlinear in both the state and observation equations. In this case, the filters with Taylor expansions can be applied for both state and parameter estimation. For parameter estimation, we can obtain more information and achieve better identification from the second observation equation. In this case, the parameters $\beta$ or $\gamma$ affect explicitly the consumption decision and can be inferred from the new observable.

**Estimation**

For parameter estimation, the task for the econometrician is to construct a quasi-likelihood function and carry out inference on $\theta \equiv (\alpha, \beta, \delta, \eta, \rho, \sigma_A, \sigma_y, \sigma_c)$. However, the construction of the quasi-likelihood requires slightly more numerical work than the previous exercises. The quasi-likelihood function $\mathcal{L}(\theta)$ is constructed as follows. Given a vector of parameter values $\theta$, I calculate the non-stochastic steady state of the system, $C_{ss}$ and $K_{ss}$, based on Eqs. (2.49)–(2.50). Given these values, I solve the model in (2.51) using perturbation methods. As a result, I obtain values for the derivatives of the policy functions $\phi$ and $\pi$. Finally, I calculate the moments $(2.56) - (2.60)$ and obtain the value of the quasi-likelihood function at $\theta$, using Algorithm (2.3.6).

**Monte Carlo Simulation Results**

In this section, I test the filter using Monte Carlo methods. First, I simulate a sample path of size $T = 500$ using the parameter values of Table 2.10. Then, using an order of $M = 4$,[33] I estimated the filtered series of $\widehat{k}_t$ and $\widehat{a}_t$ using two state-space models and the true parameter values. The first representation is in Eqs. (2.54) and the second in Eqs. (2.62). The results are shown in figure 2.8. We learn in this case, that the filtered series under both models provide similar results; indeed, the correlation coefficient between both series is above 90%. To assess the accuracy of both filters, I repeat the same exercise 500 times and keep track of the MSEs of the state variables and their simulated counterparts, as well as for the simulated GDP and consumption series. The results are displayed in Tables 2.11 and 2.12. Clearly, in this example, the addition of a second observation equation does not improve the identification of the state variables, since the average MSEs are higher in the second state-space representation.

For parameter estimation the results are different. As in the previous models, basic identification exercises are performed. Sample paths of log-investment ($\widehat{k}_t$) and random shocks ($\widehat{a}_t$) are simulated with $T = 500$ using the parameters of Table 2.10. As in the filtering exercise, the degree of approximation of the quasi-likelihood function is $M = 4$. For each simulation, the likelihood function of each of the state-space representations is evaluated numerically in the true vector of parameter values except the one shown on the $x$-axis. The results are shown in Figure 2.9. The two figures on the left represent the quasi-likelihood function of the first state-space representation, while the other figures were calculated with the second.

We learn from these exercises that the quasi-likelihood functions are continuous with respect to the parameter values. This is an important advantage with respect to particle filters, since standard numerical optimization methods can be applied for statistical inference. In addition, I analyze the identification of the subjective discount factor, $\beta$, and the capital share, $\alpha$. Flury and Shephard (2011) present evidence of misidentification of these parameter values. However, by adding a second observation equation, we learn that we can identify both parameter values, as the quasi-likelihood function presents a concave shape. Moreover, the maximum is achieved in a point very close to the true parameter value, which is represented by the vertical dotted lines.

Finally, to provide more evidence of the accuracy of QML estimates, I simulate 100 independent samples of $T = 100$ monthly returns with the parameters of Table 2.10. Based on the second-order approximations of Schmitt-Grohe and Uribe (2004), I construct a quasi-likelihood function for each of the two state-space representations with an order of approximation of $M = 4$. For each simulated series I maximize these quasi-likelihood functions using the true vector of

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

parameter values as starting point and obtain parameter estimates.[34] The results are shown in Table 2.13. The average estimate and the standard deviation of the estimates constructed with the simulated series is reported in the table. In general, the standard deviation of the parameter estimates of the second model is smaller than the standard deviation of the first set of parameter estimates. Therefore, I confirm that including consumption as a second observation equation helps to better identify the set of parameter values. Although the number of observations is relatively small, in almost all cases the true parameter value differs from the average estimate by less than two sample standard deviations. Finally, in the second model, the standard deviation of the correlation coefficient estimates is relatively high. This finding is consistent with the results of Flury and Shephard (2011), that show that the correlation coefficient estimates are not well identified. Estimating the DSGE model is done in a spirit of demonstrating the workings and capabilities of the Kalman filter with Taylor series rather than gaining any new insight on model parameters.

## 2.6 Robustness Checks

In this section I discuss robustness checks for the filtering method with Taylor series approximations. In the first robustness check, I test the filter in one state-space representation commonly used in the nonlinear filtering literature to illustrate the operation of the filter when systems are highly nonlinear. As a second robustness check, I study the effect of dimensionality in the filter, by studying a multivariate stochastic volatility model recently studied in Chib, Nardari, and Shephard (2006) and Chib, Omori, and Asai (2009).

### 2.6.1 Highly Nonlinear Systems

The first example is a benchmark commonly used in the nonlinear filtering literature and corresponds to the univariate nonstationary growth model from Gordon, Salmond, and Smith (1993). The following nonlinear model is considered:

$$x_t = 0.5x_{t-1} + \frac{25x_{t-1}}{(1+x_{t-1}^2)} + 8\cos\left(1.2\left(t-1\right)\right) + w_t, \tag{2.63}$$

and

$$y_t = \frac{x_t^2}{20} + v_t, \tag{2.64}$$

where $w_t$ and $v_t$ are zero-mean Gaussian white noise with variances 10 and 1, respectively.

[34] As a robustness check, the quasi-likelihood function was maximized numerically using different initial points randomly chosen from the parameter space. Although the numerical results are similar to the ones obtained using the true value as initial point, the correlation coefficient ($\rho$) is not well identified in a few of cases, as the estimate is a corner solution. As documented by Rytchkov (2012), this issue can be corrected by including additional observation equations or increasing the sample size.

**State-Space Representation and Implementation**

Eqs. (2.63) and (2.64) define a state-space model with state variable $x_t$. Both, the observation and transition equations are clearly nonlinear. The observation equation is quadratic in $x_t$; i.e., $h(x_t) = \frac{x_t^2}{20}$, with a random noise $v_t$ with variance $R \equiv 10$. The transition equation is nonlinear as well, defined by the function

$$g(x) = 0.5x + \frac{25x}{(1+x^2)}, \tag{2.65}$$

with variance $Q \equiv 1$. The deterministic function in Eq. (2.63) is measurable at time $t$ and is incorporated in the filtering exercise as a known constant at the beginning of each time step.

To apply the Taylor series approximation, it is necessary to obtain the Taylor series expansion of the functions $h$ and $g$. However, a major complication occurs with the Taylor series of $g$, since its Taylor series does not converge uniformly in the entire domain. As a result, Eq. (2.12) does not hold. An alternative approach consists of applying Lemma 2.6.1; and obtain a power series functions $x/(1+x^2), x/(1+x^2)^2$ and $x^2/(1+x^2)^2$ in terms of $\frac{\lambda-1-x^2}{\lambda}$. With this new representation, we can estimate the mean, variance and covariance necessary for the filtering recursions.

**Lemma 2.6.1** *Let $x \in \mathbb{R}$, $|x| < \sqrt{2\lambda - 1}$ where $\lambda > \frac{1}{2}$; then*

$$\frac{x}{(1+x^2)} = \sum_{j=0}^{\infty} \frac{x}{\lambda} \left( \frac{\lambda - 1 - x^2}{\lambda} \right)^j, \tag{2.66}$$

$$\frac{x^2}{(1+x^2)} = \sum_{j=0}^{\infty} \frac{x^2}{\lambda} \left( \frac{\lambda - 1 - x^2}{\lambda} \right)^j \tag{2.67}$$

*and*

$$\frac{x^2}{(1+x^2)^2} = \sum_{j=0}^{\infty} j \left( \frac{x}{\lambda} \right)^2 \left( \frac{\lambda - 1 - x^2}{\lambda} \right)^{j-1}. \tag{2.68}$$

***Proof*** *See Appendix A.2.*

From Lemma 2.6.1, we learn that to estimate the mean and variance of Eq. (2.65) we need to estimate moments of the form

$$\mathbb{E}\left[ \frac{x}{\lambda} \left( \frac{\lambda - 1 - x^2}{\lambda} \right)^j \right], \ j = 0, 1, 2, ...M$$

$$\mathbb{E}\left[ \frac{x^2}{\lambda} \left( \frac{\lambda - 1 - x^2}{\lambda} \right)^j \right], \ j = 0, 1, 2, ...M$$

and

$$\mathbb{E}\left[\left(\frac{x}{\lambda}\right)^2\left(\frac{\lambda-1-x^2}{\lambda}\right)^{j-1}\right],\ j=1,2,...M$$

These moments can be calculated directly from the results of the first section for an order of $M = 20$. Once the moments are calculated, $\lambda$ has to be large enough to guarantee that the estimation results converge.

**Monte Carlo Simulation Results**

Following Gordon, Salmond, and Smith (1993), I assume that the initial state is $x_0 = 0.1$ and simulate a realization of Eq. (2.63) of 50 time steps. The filters are initialized with a prior distribution $\mathcal{N}(0, 2)$. Figure 2.10a compares the filter with Taylor series approximations with the true state estimates. The true state is represented by a star, the mean of the filter with Taylor series is given as a solid line and the dashed lines give the 95% probability interval.[35] In this case, none of the simulations exceeded the 95% intervals. Figure 2.10b compares the same simulated series with the results obtained from the filter with Taylor series as well as the ones obtained with the particle filter constructed with 1000 particles. The solid line gives the filtered series based on the Taylor approximations, while the dashed line shows the estimate of the particle filter. For this specific case, the correlation coefficient between the filtered series is above 90%.

### 2.6.2 Multivariate Stochastic Volatility Models

As a robustness check to high-dimensional systems, I consider a multivariate stochastic volatility model.[36] Consider a $p \times 1$ vector of asset log-returns $y_t = (y_{1t}, ..., y_{pt})^\top$ with a constant mean vector $\mu = (\mu_1, ..., \mu_p)^\top$ and stochastic time-varying variance covariance matrix $V_t$. More specifically, the vector $y_t$ is specified as

$$y_t \,|h_t \sim \mathcal{N}(\mu, V_t), \qquad t = 1, ..., n, \tag{2.69}$$

where $h_t$ is a scalar or a vector stochastic process. The variance matrix $V_t$ is a function of $h_t$. In this case, the latent variables $h_t$ are modeled as an autoregressive process of order one. It is also assumed that the variance–covariance matrix can be decomposed by $V_t = D\Sigma_t^2 D^\top$, where the matrix $D$ is a lower unity triangular matrix. This is the Cholesky decomposition with a diagonal matrix $\Sigma_t^2$ that is time-varying according to the stochastic process $h_t$. In this case both the variances and correlations implied by $V_t$ are time-varying. The resulting multivariate $SV$

[35] This interval is constructed using the 2.5 and 97.5 percentiles of the normal distribution.

[36] See Chib, Omori, and Asai (2009) and Asai, McAleer, and Yu (2006) for comprehensive reviews of multivariate stochastic volatility models.

model based on this decomposition is given by

$$\begin{aligned} y_t &= \mu + D\Sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, I), \\ h_{t+1} &= \gamma + \Phi h_t + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \Sigma_\eta), \end{aligned} \tag{2.70}$$

where $\Sigma_t = \exp\left(\frac{1}{2} diag\left(h_t\right)\right)$ and $D$ can be regarded as a nonsingular load matrix. The model is a special case of the class of multivariate SV models that was originally proposed by Harvey, Ruiz, and Shephard (1994) and Shephard (1996) and further extended by Chib, Nardari, and Shephard (2006).

**State-Space Representation and Implementation**

For easiness of exposition, I assume that $y_t$ is a two-dimensional vector of demeaned returns.[37] As a result, the representation is of the form

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} D_{11} & 0 \\ D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} \exp\left(\frac{h_{1t}}{2}\right)\varepsilon_{1t} \\ \exp\left(\frac{h_{2t}}{2}\right)\varepsilon_{2t} \end{bmatrix}, \qquad \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \sim \mathcal{N}(0, I_2), \tag{2.71}$$

where

$$\begin{bmatrix} h_{1t+1} \\ h_{2t+1} \end{bmatrix} = \begin{bmatrix} \Phi_{11} & 0 \\ 0 & \Phi_{22} \end{bmatrix} \begin{bmatrix} h_{1t} \\ h_{2t} \end{bmatrix} + \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix}, \qquad \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} \sim \mathcal{N}(0, \Sigma_\eta). \tag{2.72}$$

For simplification purposes, I represent the system in Eqs. (2.71) and (2.72) in terms of the vector of state variables, $x_t = [x_{1t}, x_{2t}, x_{3t}, x_{4t}, x_{5t}, x_{6t}]^\top = [h_{1t}, \eta_{1t}, \varepsilon_{1t}, h_{2t}, \eta_{2t}, \varepsilon_{2t}]$. Therefore, the observation equations become

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} D_{11} & 0 \\ D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} \exp\left(\frac{x_{1t}}{2}\right) x_{3t} \\ \exp\left(\frac{x_{4t}}{2}\right) x_{6t} \end{bmatrix}. \tag{2.73}$$

The four-dimensional vector of state variables evolves according to the following:

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \\ x_{4t} \\ x_{5t} \\ x_{6t} \end{bmatrix} = \begin{bmatrix} \Phi_{11} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Phi_{22} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \\ x_{4t-1} \\ x_{5t-1} \\ x_{6t-1} \end{bmatrix} + \Gamma \begin{bmatrix} \eta_{1t} \\ \varepsilon_{1t} \\ \eta_{2t} \\ \varepsilon_{2t} \end{bmatrix}, \qquad \begin{bmatrix} \eta_{1t} \\ \varepsilon_{1t} \\ \eta_{2t} \\ \varepsilon_{2t} \end{bmatrix} \sim \mathcal{N}(0, Q), \tag{2.74}$$

where

[37] An observational equivalent model is obtained by setting $\gamma = 0$ and $D$ as a lower triangular matrix with nonzero values on its leading diagonal.

$$Q = \begin{bmatrix} \sigma_{11} & 0 & \sigma_{12} & 0 \\ 0 & 1 & 0 & 0 \\ \sigma_{12} & 0 & \sigma_{22} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\Gamma^{\top} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Clearly, from Eq. (2.74), we learn that the system has a linear transition equation; i.e., $g(x_t) = \Phi \cdot x_t$, with covariance matrix $Q$. In this case, the noise vector of the observation equation is assumed to be identically zero; therefore, the covariance of the observation matrix is zero (i.e., $R \equiv 0$). However, this representation is flexible enough and can be extended to a more general correlation structure between the innovations of the state and transition equations. The vector of observation equations, $h(x_t) = [h_1(x_t), h_2(x_t)]^{\top}$, is represented by Eqs. (2.73). Although, the moments involved in the filtering recursions have closed-form expressions, the Taylor approximations can be implemented.

As in the one-dimensional case, by modeling the product of the observation Eqs. (2.73), we can perform better moment identification. As a result, I include three more observation equations to the vector of observables:

$$\begin{aligned} y_{3t} &= y_{1t}^2 = D_{11}^2 e^{2x_{1t}} x_{3t}^2, \\ y_{4t} &= y_{2t}^2 = \left(D_{21} e^{x_{1t}} x_{3t} + D_{22} e^{x_{2t}} x_{4t}\right)^2, \\ y_{5t} &= y_{1t} \cdot y_{2t} = D_{11}\left(D_{21} e^{2x_{1t}} x_{3t}^2 + D_{22} e^{x_{1t}+x_{2t}} x_{3t} x_{4t}\right). \end{aligned} \tag{2.75}$$

The first two equations in 2.75 follow from the one-dimensional case. Moreover, I include the third observation equation for robustness. The overall system has five nonlinear observation equations, given by Eqs. (2.73) and (2.75), and six state variables that satisfy the law of motion in Eq. (2.74). The filtering exercise is performed with an order of approximation $M = 8$. However, closed-form expressions for means and covariances can be calculated directly by applying the results shown in Appendix 2.3.2.

**Monte Carlo Simulation Results**

To test the quality of the approximation, I simulate a random path of log-volatilities and shocks with $T = 100$, using the parameters from Jungbacker and Koopman (2006) presented in Table 2.14. The results are shown in Figures 2.11a and 2.11b. We learn from this implementation that the filter with Taylor series provides similar estimates to the particle filter. Surprisingly,

the filter with Taylor series provides results as accurate as the standard particle filter. For this random simulation, the correlation between the filtered series with Taylor expansions and the ones with particle filters is above 90%.

## 2.7 Concluding Remarks

In this chapter, I propose a nonlinear filter based on Taylor series approximations, which is based on efficient calculation of derivatives and can be applied for state and parameter estimation. My results suggest that filtering methods via the Taylor series filter are superior to conventional methods such as the EKF or the UKF. I also find that the filter with Taylor approximations is as accurate as the standard particle filter and at least twenty times faster than conventional Monte Carlo filters.

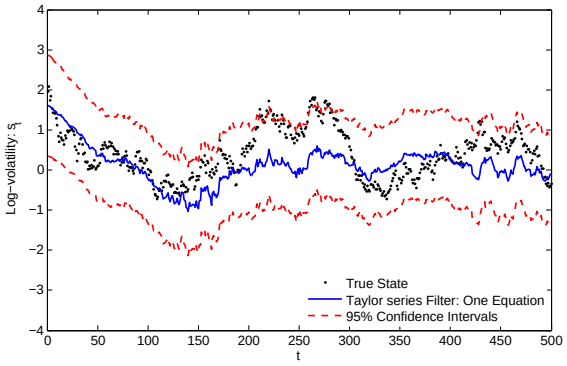
I test the filter in a number of models such as univariate and multivariate stochastic volatility models, a risk-return model, a dynamic stochastic general equilibrium model and a highly nonlinear filter. My findings suggest that the filter with Taylor approximations provides accurate results for both state and parameter estimation. The filter provides accurate estimates in high-dimensional and nonlinear systems. The time efficiency is comparable to the low-dimensional case, in which some applications are at least twenty times faster than standard particle filters.

I also find that by adding more observation equations to the state-space model, such as the squared value of the current observation equation, the unobserved states can be better identified. With this augmented state-space model, the filtered estimates are comparable with those obtained via standard particle filters. The filter with Taylor series can be applied for inference purposes since a quasi-likelihood function is obtained with the filtering recursions. In all the examples in this chapter, I find continuous quasi-likelihood functions with respect to the parameter values and conventional methods for numerical optimization can be applied for their estimation. The efficiency of the state estimation can be used for parameter estimation.

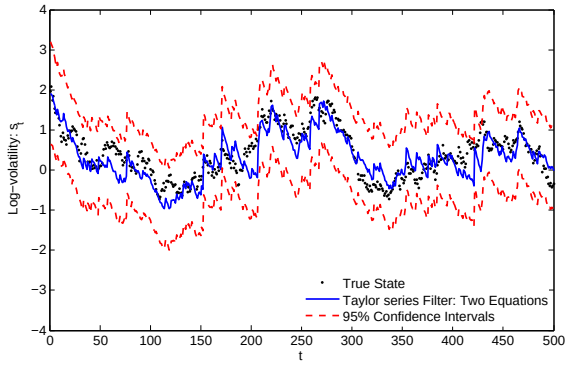
Finally, the filter with Taylor series approximations can be applied to a more general class of models by combining its analytical tractability with Monte Carlo methods, via Rao-Blackwellised filters. By combining these techniques, the density of the filter with Taylor approximations can provide accurate estimates of the true density of the filter, up to a normalization constant.

Although the filters provide accurate results for a number problems, some care should be taken in the modeling and implementation. Some limitations arise when the function to be approximated is non-differentiable or the Taylor series approximations are not uniformly convergent. One approach to circumvent this issue is by taking the Taylor series in another center of expansion or by changing the scale of the state variables. The results can be extended to non-differentiable functions, as long as the function can be approximated with a power series. Another shortcoming is that significant work has to be done to include the correct observation equation within the estimation process.

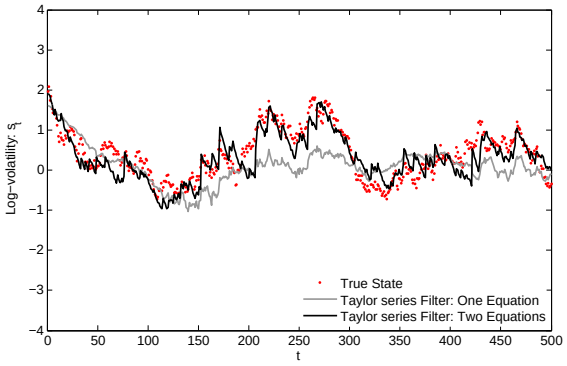
## 2.8 Figures and Tables



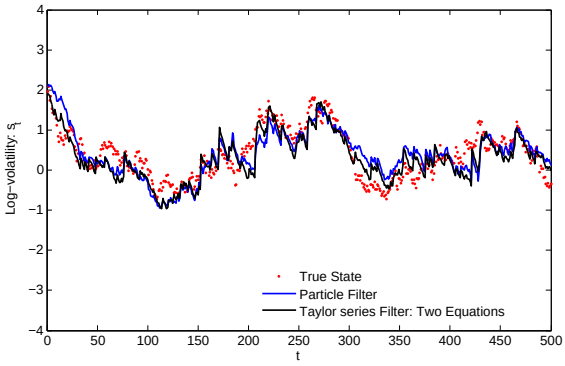

**(a)** Stochastic Volatility Model 1




**(b)** Stochastic Volatility Model 2

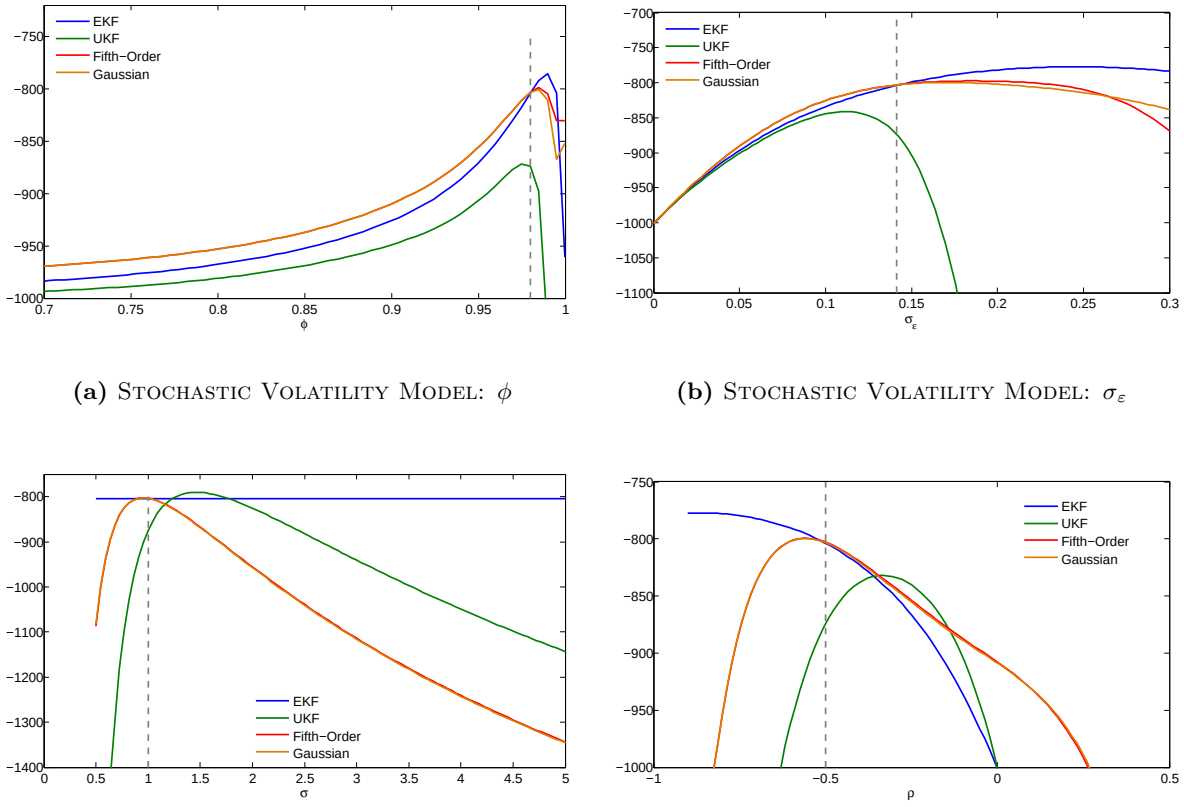



**(c)** Stochastic Volatility Models 1 and 2




**(d)** Stochastic Volatility Model 2: Comparison

**Figure 2.1.** Stochastic Volatility Model: Filter Performance.
This figure compares a simulated time series of 500 observations for the standard stochastic volatility model with its filtered values. The filtered estimates in the two figures on top were calculated using a fifth-order of 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 $\rho = -0.5$.



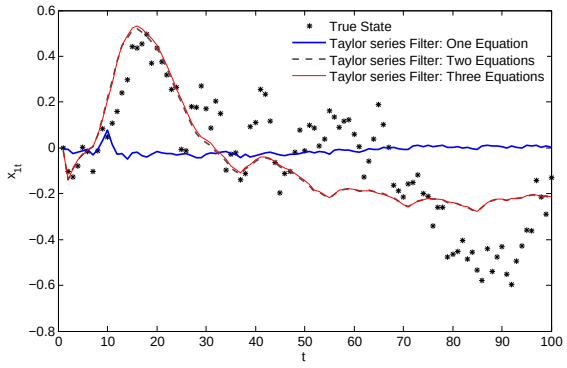

**(a)** Stochastic Volatility Model: $\phi$

**(b)** Stochastic Volatility Model: $\sigma_\varepsilon$

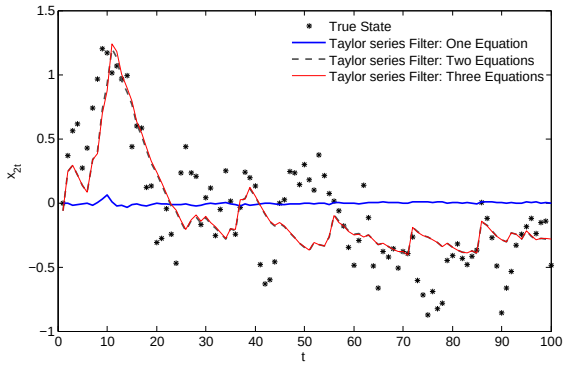
**(c)** Stochastic Volatility Model: $\overline{\sigma}$

**(d)** Stochastic Volatility Model: $\rho$

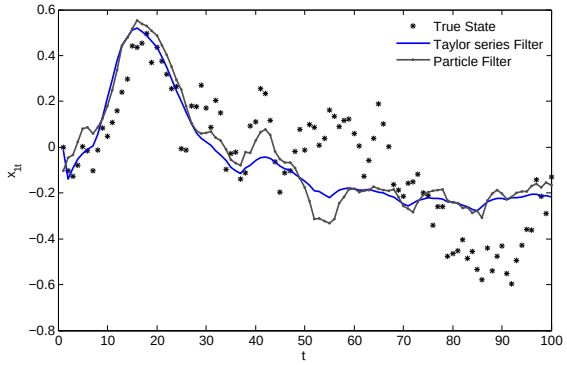
**Figure 2.2.** Stochastic Volatility Model: Quasi-Likelihood Contours
This figure plots the quasi-likelihood function of a standard stochastic volatility model for different sets of parameter vectors. The plots show the quasi-log-likelihood function of the data for different values of $\phi$ (top left), $\sigma_\varepsilon$ (top right), $\overline{\sigma}$ (bottom right) and $\rho$ (bottom left). The vertical dashed lines represent the parameter values that were used to simulate the data.



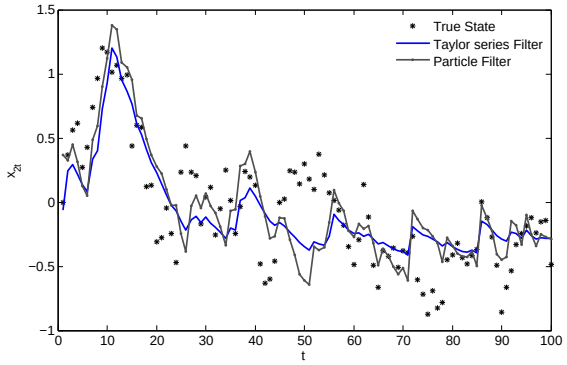

**(a)** Different Observation Equations: log-Expected Returns




**(b)** Different Observation Equations: log-Volatility




**(c)** Risk-Return Model: log-Expected Returns




**(d)** Risk-Return Model: log-Volatility

**Figure 2.3.** Risk-Return Model: Filter Performance

This figure plots the filtered series of a random draw of $T = 100$ returns simulated from the model by Brandt and Kang (2004). The parameters used for the simulation are $a_{11} = 0.8589$, $a_{21} = -0.0531$, $a_{12} = 0.1081$, $a_{22} = 0.9237$,$b_{11} = 0.0076$,$b_{22} = 0.0554$ and $\rho = -0.6345$.

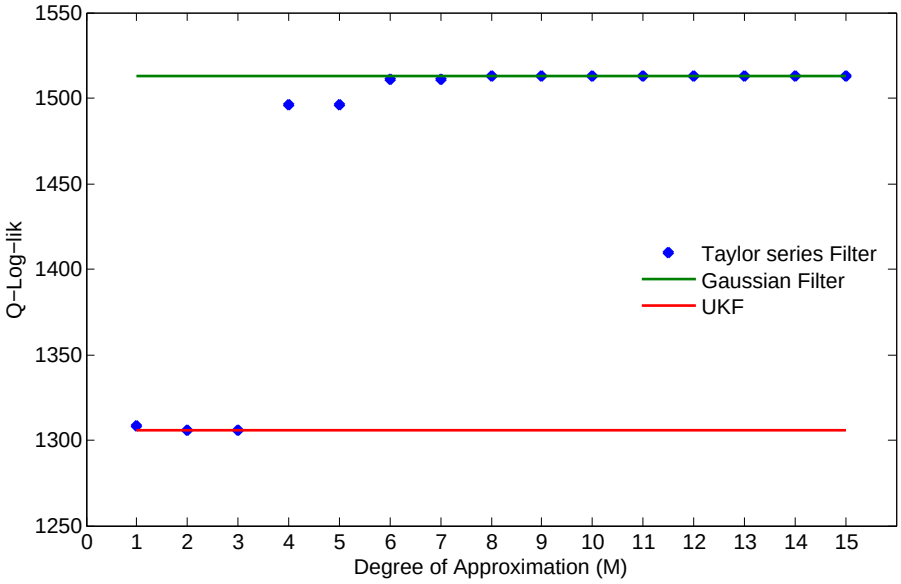
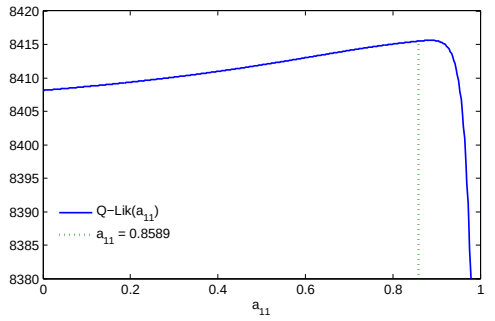
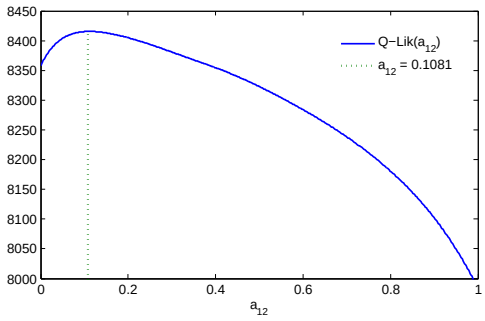



**Figure 2.4.** Risk-Return Model: Order of Approximation

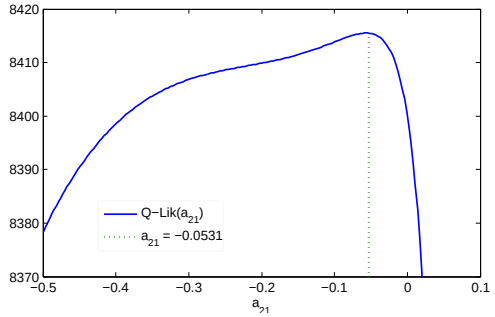
This figure plots the quasi-likelihood function of a random simulation of the model by Brandt and Kang (2004) evaluated with the parameter values $a_{11} = 0.8589$, $a_{21} = -0.0531$, $a_{12} = 0.1081$, $a_{22} = 0.9237$, $b_{11} = 0.0076$, $b_{22} = 0.0554$ and $\rho = -0.6345$. The plot displays the values of the quasi-likelihood function for different orders of approximation, $M = 1, 2, ..., 15$ (asterisks). The continuous lines represent the quasi-likelihood functions constructed with the UKF and the Gaussian filters.



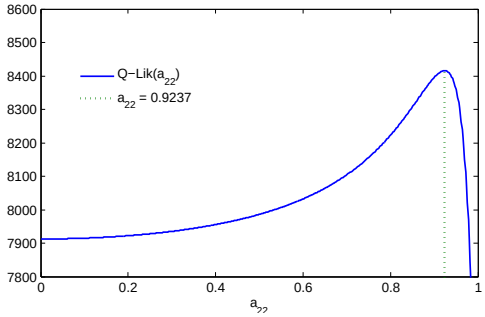

**(a)** Risk-Return Model: $a_{11}$




**(b)** Risk-Return Model: $a_{12}$



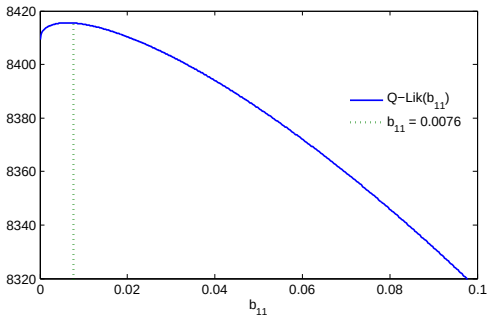

**(c)** Risk-Return Model: $a_{21}$



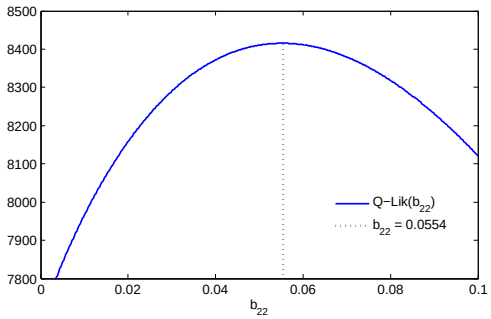

**(d)** Risk-Return Model: $a_{22}$

**Figure 2.5.** Risk-Return Model: Quasi-Likelihood Contours

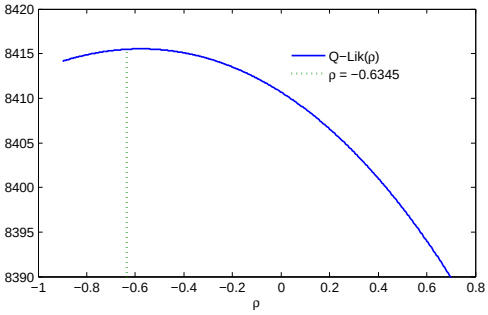
This figure plots the quasi-likelihood function of a random draw of $T = 5000$ returns simulated from the model by Brandt and Kang (2004). The parameters used for the simulation are $a_{11} = 0.8589$, $a_{21} = -0.0531$, $a_{12} = 0.1081$, $a_{22} = 0.9237$, $b_{11} = 0.0076$, $b_{22} = 0.0554$ and $\rho = -0.6345$.




**(a)** Risk-Return Model: $b_{11}$

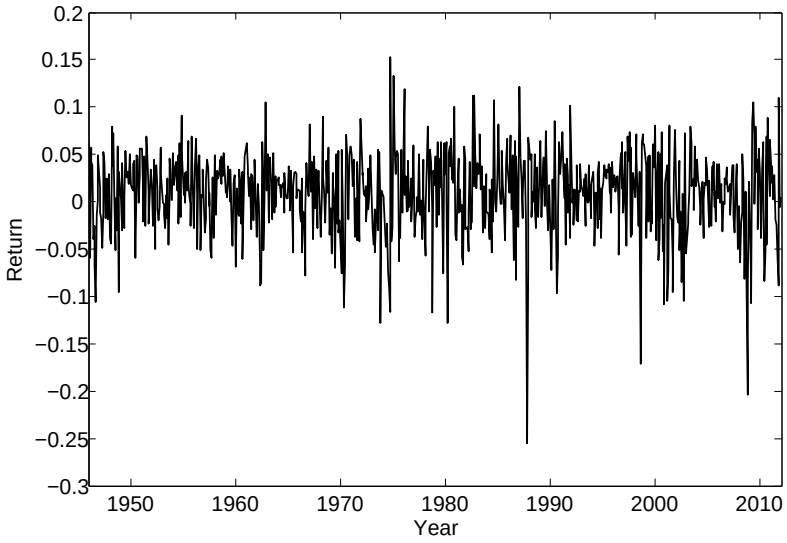



**(b)** Risk-Return Model: $b_{22}$

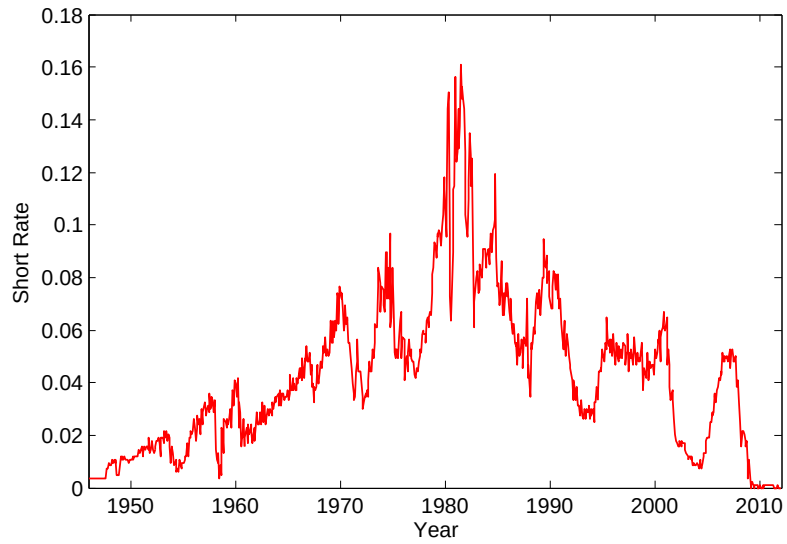



**(c)** Risk-Return Model: $\rho$

**Figure 2.6.** Risk-Return Model: Quasi-Likelihood Contours (Cont.)
This figure plots the quasi-likelihood function of a random draw of $T = 5000$ returns simulated from the model by Brandt and Kang (2004). The parameters used for the simulation are $a_{11} = 0.8589$, $a_{21} = -0.0531$, $a_{12} = 0.1081$, $a_{22} = 0.9237$, $b_{11} = 0.0076$, $b_{22} = 0.0554$ and $\rho = -0.6345$.



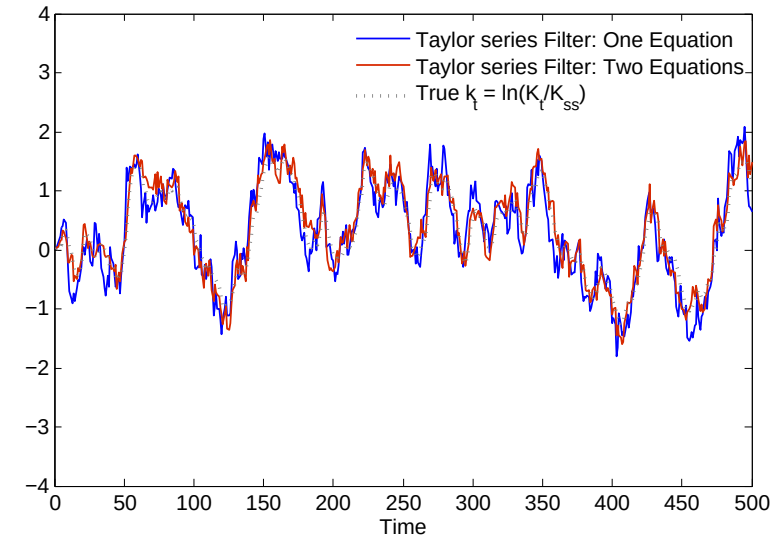

**(a)** Market Returns



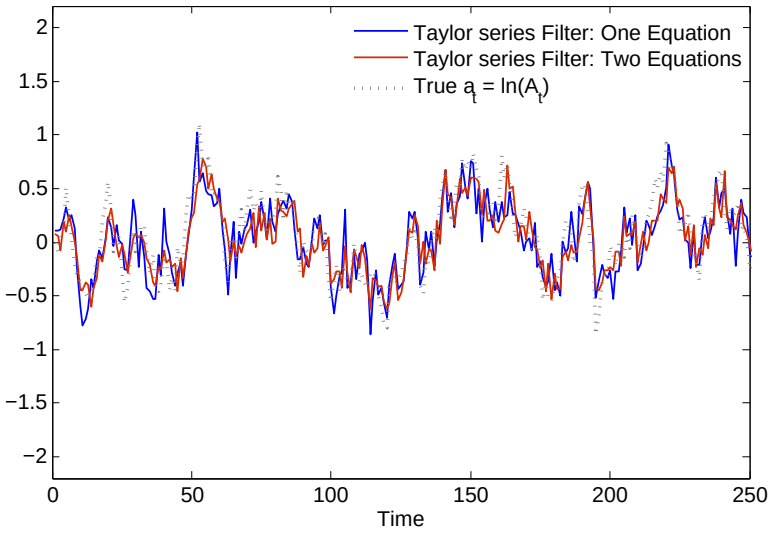

**(b)** Short Rate

**Figure 2.7.** Risk-Return Model: Data

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



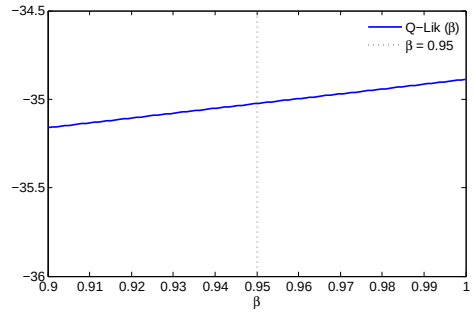

**(a)** Filtered Log-Investment $\left(k_t = \ln\left(\frac{K_t}{K_{ss}}\right)\right)$



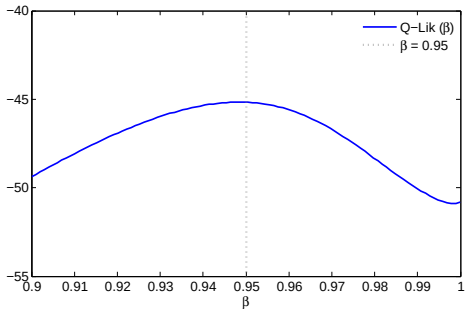

**(b)** Filtered Log-TFP Shock $(a_t = \ln(A_t))$

**Figure 2.8.** DSGE Model: Filter Performance

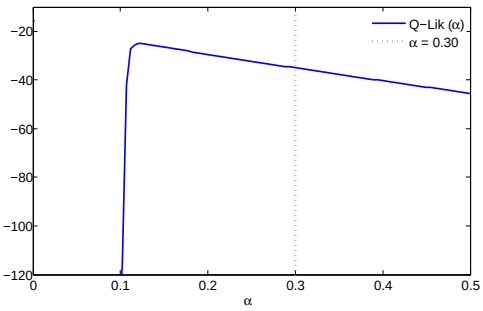
This figure plots the filtered estimates of the state variables evaluated in a simulated sample path of size $T = 500$ with the parameter values $\beta = 0.95$, $\delta = 0.15$, $\alpha = 0.30$, $\rho = 0.90$,$\gamma = 3$,$\sigma_y = 0.30$ and $\sigma_\varepsilon = 0.2$.



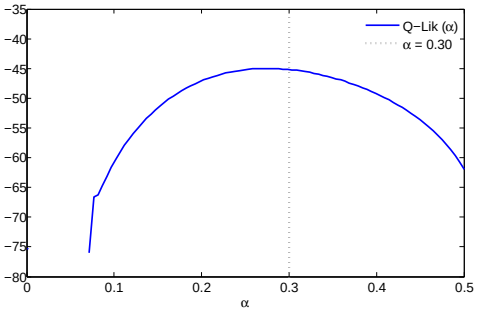

**(a)** QL function $\beta$: Model 1




**(b)** QL function $\beta$: Model 2



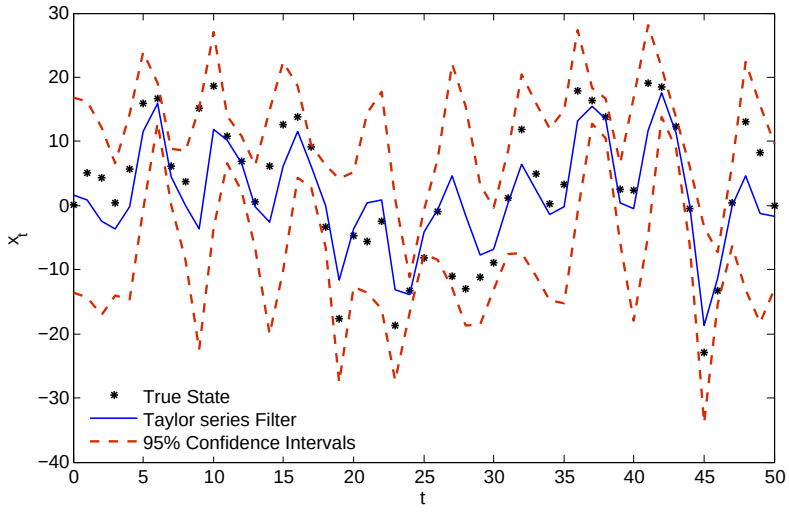

**(c)** QL function $\alpha$: Model 1



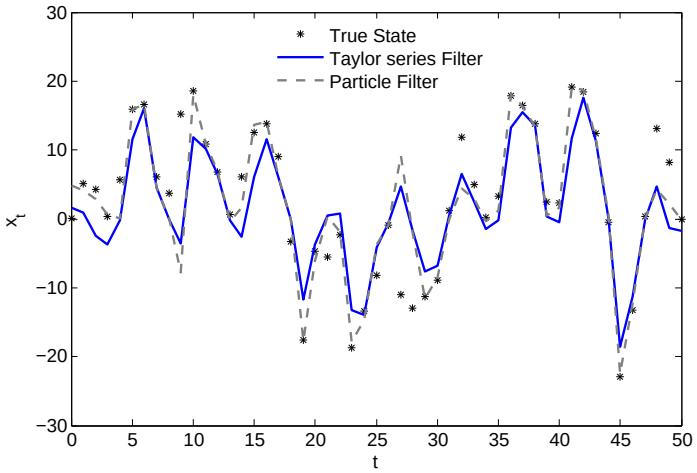

**(d)** QL function $\alpha$: Model 2

**Figure 2.9.** DSGE Model: Quasi-Likelihood Contours

This figure plots the quasi-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 $\sigma_\varepsilon = 0.2$.

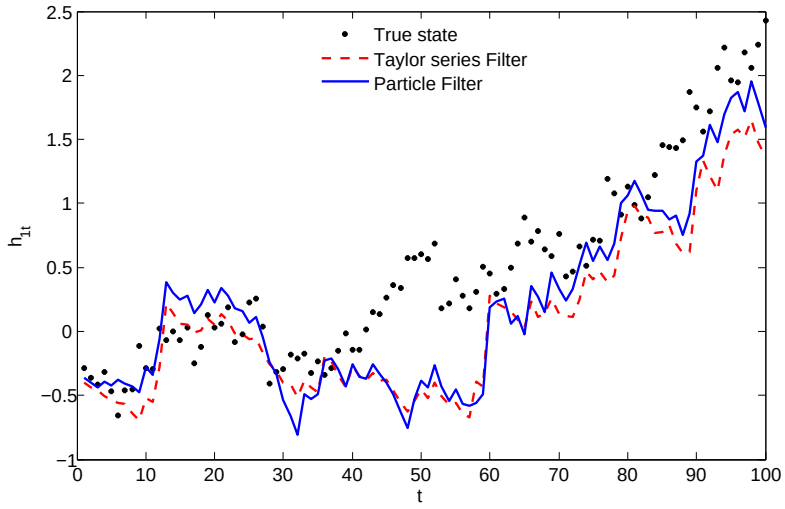



**(a)** Filter Performance

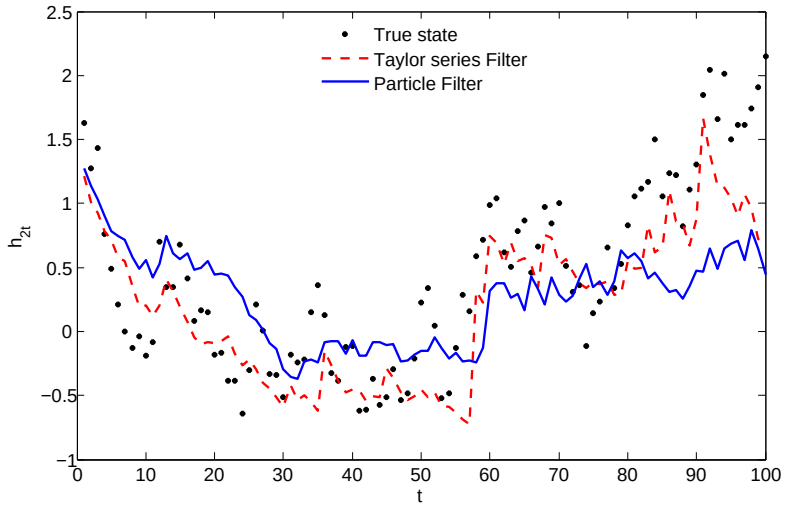



**(b)** Filter Comparison

**Figure 2.10.** Nonlinear Model: Filter Performance
This figure plots the filtered series of the nonlinear model by Gordon, Salmond, and Smith (1993).




**(a)** Filter Performance: first log-volatility process ($h_{1,t}$)




**(b)** Filter Performance: second log-volatility process ($h_{2,t}$)

**Figure 2.11.** Multivariate Stochastic Volatility Model: Filter Performance
This figure plots the filtered series of the multivariate version of the stochastic volatility model by Chib, Nardari, and Shephard (2006).

**Table 2.1.** Stochastic Volatility Model (One Observation Equation): Simulation Results

| **Filtering Method** | **MSE** $(\log \sigma_t^2)$ | | **MSE** $(y_t)$ | |
|---|---|---|---|---|
| | **Average** | **Std. Dev.** | **Average** | **Std. Dev.** |
| EKF | 0.3450 | 0.1529 | 1.3100 | 0.4688 |
| UKF | 0.3430 | 0.1406 | 1.3122 | 0.4626 |
| TKF (2) | 0.3321 | 0.1289 | 1.3108 | 0.4630 |
| TKF (3) | 0.3321 | 0.1289 | 1.3108 | 0.4630 |
| TKF (4) | 0.3366 | 0.1400 | 1.3111 | 0.4630 |
| TKF (5) | 0.3366 | 0.1400 | 1.3111 | 0.4630 |
| TKF (6) | 0.3412 | 0.1436 | 1.3111 | 0.4630 |
| TKF (7) | 0.3412 | 0.1436 | 1.3111 | 0.4630 |
| TKF (8) | 0.3424 | 0.1444 | 1.3111 | 0.4630 |
| TKF (9) | 0.3424 | 0.1444 | 1.3111 | 0.4630 |
| TKF (10) | 0.3426 | 0.1445 | 1.3111 | 0.4630 |
| TKF (11) | 0.3426 | 0.1445 | 1.3111 | 0.4630 |
| TKF (12) | 0.3426 | 0.1445 | 1.3111 | 0.4630 |
| Gaussian Filter | 0.3426 | 0.1445 | 1.3111 | 0.4630 |
| Particle Filter | 0.1691 | 0.0416 | 0.7439 | 0.2597 |

**Monte Carlo simulation results.** This table presents the mean and variance of the MSEs between the simulated and the filtered values of the log-volatility $(\log \sigma_t^2)$ process and the observable $(y_t)$ of the standard stochastic volatility model:

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

The results are based on 500 independent samples of $T = 500$ simulated from the model with the parameters $\phi = 0.98$, $\sigma_\varepsilon = 0.1414$, $\overline{\sigma} = 1$ and $\rho = -0.5$ .

**Table 2.2.** Stochastic Volatility Model (Two Observation Equations): Simulation Results

| Filtering Method | MSE $(\log \sigma_t^2)$ | | MSE $(y_t)$ | |
|---|---|---|---|---|
| | **Average** | **Std. Dev.** | **Average** | **Std. Dev.** |
| EKF | 0.3649 | 0.1654 | 1.2693 | 0.4077 |
| UKF | 0.2675 | 0.1022 | 1.3073 | 0.4651 |
| TKF (2) | 0.6084 | 0.4301 | 1.3314 | 0.5509 |
| TKF (3) | 0.6084 | 0.4301 | 1.3314 | 0.5509 |
| TKF (4) | 0.4047 | 0.3116 | 1.3120 | 0.4718 |
| TKF (5) | 0.4047 | 0.3116 | 1.3120 | 0.4718 |
| TKF (6) | 0.1805 | 0.0450 | 1.3073 | 0.4647 |
| TKF (7) | 0.1805 | 0.0450 | 1.3073 | 0.4647 |
| TKF (8) | 0.1658 | 0.0376 | 1.3073 | 0.4647 |
| TKF (9) | 0.1658 | 0.0376 | 1.3073 | 0.4647 |
| TKF (10) | 0.1636 | 0.0367 | 1.3073 | 0.4647 |
| TKF (11) | 0.1636 | 0.0367 | 1.3073 | 0.4647 |
| TKF (12) | 0.1632 | 0.0365 | 1.3073 | 0.4647 |
| Gaussian Filter | 0.1632 | 0.0365 | 1.3073 | 0.4647 |
| Particle Filter | 0.1691 | 0.0416 | 0.7439 | 0.2597 |

**Monte Carlo simulation results.** This table presents the mean and variance of the MSEs between the simulated and the filtered values of the log-volatility $(\log \sigma_t^2)$ process and the observables $(y_{1,t}, y_{2,t})$ of the standard stochastic volatility model:

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

The results are based on 500 independent samples of $T = 500$ simulated from the model with the parameters $\phi = 0.98$, $\sigma_\varepsilon = 0.1414$, $\overline{\sigma} = 1$ and $\rho = -0.5$ .

**Table 2.3.** Stochastic Volatility Model: CPU Time

| Filtering Method | Model 1 | | Model 2 | |
|---|---|---|---|---|
| | Average | Std. Dev. | Average | Std. Dev. |
| EKF | 0.0292 | 0.0049 | 0.0827 | 0.0126 |
| UKF | 0.0848 | 0.0076 | 0.1110 | 0.0075 |
| TKF (2) | 0.1664 | 0.0152 | 0.3234 | 0.0221 |
| TKF (3) | 0.2577 | 0.0246 | 0.4306 | 0.0342 |
| TKF (4) | 0.3804 | 0.0209 | 0.5483 | 0.0292 |
| TKF (5) | 0.6027 | 0.0785 | 0.7828 | 0.1000 |
| TKF (6) | 0.8498 | 0.0650 | 1.0252 | 0.0650 |
| TKF (7) | 1.1805 | 0.0519 | 1.3624 | 0.0534 |
| TKF (8) | 1.6703 | 0.1242 | 1.8646 | 0.1449 |
| TKF (9) | 2.6574 | 0.2697 | 2.9169 | 0.3152 |
| TKF (10) | 3.1050 | 0.2253 | 3.3022 | 0.2256 |
| TKF (11) | 3.5978 | 0.0939 | 3.7845 | 0.0983 |
| TKF (12) | 4.4738 | 0.0819 | 4.6712 | 0.0888 |
| Gaussian Filter | 0.0318 | 0.0028 | 0.0412 | 0.0037 |
| Particle Filter | 17.8776 | 0.2131 | 17.8776 | 0.2131 |

**CPU time.** This table presents the CPU time that a standard computer takes in seconds to compute the filtered values of the log-volatility ($\log \sigma_t^2$) process. The second and third columns contain the mean and standard deviation of the CPU time in seconds for a model with one observable. The fourth and fifth column are the mean and standard deviation of the CPU time in seconds for a model with two observables. The results are based on 500 independent samples of $T = 500$ simulated from the model with the parameters $\phi = 0.98$, $\sigma_\varepsilon = 0.1414$, $\overline{\sigma} = 1$ and $\rho = -0.5$ .

**Table 2.4.** Stochastic Volatility Model: Quasi-Maximum Likelihood Estimation Results

| Filtering Method | Parameters | | | |
|---|---|---|---|---|
| | $\phi$ | $\sigma_\varepsilon$ | $\overline{\sigma}$ | $\rho$ |
| | 0.9800 | 0.1414 | 1.0000 | -0.5000 |
| EKF | 0.9348 | 0.1491 | 1.2402 | -0.3598 |
| | (0.1454) | (0.0701) | (0.4162) | (0.4479) |
| UKF | 0.9287 | 0.1614 | 1.2926 | -0.4383 |
| | (0.1258) | (0.1050) | (0.4539) | (0.4656) |
| TKF (3) | 0.9241 | 0.1601 | 1.2926 | -0.4805 |
| | (0.1698) | (0.1012) | (0.4277) | (0.4663) |
| TKF (5) | 0.9292 | 0.1753 | 0.966 | -0.4663 |
| | (0.2085) | (0.1030) | (0.3975) | (0.4353) |
| TKF (7) | 0.9564 | 0.1781 | 0.8502 | -0.4852 |
| | (0.0693) | (0.1035) | (0.4103) | (0.4091) |
| TKF (9) | 0.9405 | 0.1907 | 0.8358 | -0.4418 |
| | (0.1629) | (0.1013) | (0.5483) | (0.3999) |
| Gaussian Filter | 0.9550 | 0.1542 | 0.9147 | -0.5393 |
| | (0.0820) | (0.0899) | (0.4263) | (0.3840) |

**Finite sample properties of the QML estimator.** This table presents the sample mean and standard deviation in parentheses of the QML estimates of the model:

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

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

**Table 2.5.** Stochastic Volatility Model: 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 |

**Descriptive statistics.** 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. The series was obtained from the *Real-Time Data Set for Macroeconomists* from the Federal Reserve Bank of Philadelphia.

**Table 2.6.** Stochastic Volatility Model: Parameter Estimates

| Filtering Method | Parameters | | | | |
|---|---|---|---|---|---|
| | $\phi$ | $\sigma_\varepsilon$ | $\overline{\sigma}$ | $\rho$ | $\mu_C$ |
| EKF | 0.9575 | 0.2997 | 0.1591 | 0.9991 | 0.0014 |
| | (0.0862) | (0.8654) | (0.2196) | (2.5046) | (0.0002) |
| UKF | 0.9623 | 0.0198 | 0.0035 | 0.9983 | 0.0012 |
| | (0.0616) | (0.0508) | (0.0012) | (0.0004) | (0.0027) |
| TKF (3) | 0.9623 | 0.0221 | 0.0035 | 0.8923 | 0.0012 |
| | (0.0287) | (0.0218) | (0.0007) | (0.0005) | (0.0012) |
| TKF (5) | 0.9626 | 0.0266 | 0.0035 | 0.7436 | 0.0012 |
| | (0.0724) | (0.0465) | (0.0006) | (0.0007) | (0.0014) |
| TKF (7) | 0.9627 | 0.0216 | 0.0035 | 0.9179 | 0.0012 |
| | (0.0203) | (0.0332) | (0.0006) | (0.0045) | (0.0018) |
| TKF (9) | 0.9627 | 0.0228 | 0.0035 | 0.8683 | 0.0012 |
| | (0.0446) | (0.0297) | (0.0005) | (0.0003) | (0.0011) |
| TKF (11) | 0.9626 | 0.0336 | 0.0035 | 0.5867 | 0.0012 |
| | (0.0204) | (0.0146) | (0.0002) | (0.0008) | (0.0002) |
| Gaussian Filter | 0.9626 | 0.0353 | 0.0035 | 0.5574 | 0.0012 |
| | (0.0205) | (0.0160) | (0.0001) | (0.0006) | (0.0002) |

**Estimation results.** This table presents the QML estimates of the model

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

The estimates are for monthly real consumption growth on the monthly 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 in parentheses.

**Table 2.7.** Risk-Return Model: Quasi-Maximum Likelihood Estimation Results

| 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 | 0.0112 |
| $\rho$ | -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 |
| $\rho_\mu$ | - | - | - | -0.0866 | 0.1154 | 0.7475 |
| $\rho_\sigma$ | - | - | - | - | - | - |
| **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 |
| $\rho$ | -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 |
| $\rho_\mu$ | - | - | - | -0.0517 | 0.1438 | 0.7477 |
| $\rho_\sigma$ | -0.2541 | -0.6629 | 0.3716 | -0.2430 | -0.5841 | 0.4675 |

**Estimation results.** This table describes the sampling distribution of the QML estimates of the model:

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

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

where

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

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

**Table 2.8.** Risk-Return Model: 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 |

**Descriptive statistics.** 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 2.9.** Risk-Return Model: Parameter Estimates

| Parameters | Model A | | Model B | |
|---|---|---|---|---|
| | **Estimate** | **Std. Error** | **Estimate** | **Std. Error** |
| $a_{11}$ | 0.9436 | 0.0662 | 0.9586 | 0.0161 |
| $a_{21}$ | -0.0778 | 0.1146 | -0.0285 | 0.0065 |
| $a_{12}$ | 0.3799 | 0.1171 | 0.3054 | 0.0253 |
| $a_{22}$ | 0.7317 | 0.1298 | 0.8750 | 0.0001 |
| $b_{11}$ | 0.1684 | 0.0296 | 0.1739 | 0.0015 |
| $b_{22}$ | 0.0002 | 0.0316 | 0.0024 | 0.0015 |
| $\rho$ | -0.1306 | 0.0091 | -0.9000 | 0.0014 |
| $\overline{\mu}$ | 0.0047 | 0.0165 | 0.0047 | 0.0012 |
| $\overline{\sigma}$ | 0.0434 | 0.0070 | 0.0437 | 0.0002 |
| $\rho_\mu$ | - | - | -0.6642 | 0.0170 |
| $\rho_\sigma$ | - | - | - | - |
| $\mathcal{L}$ | 1366.23 | | 1373.35 | |
| **Parameters** | **Model C** | | **Model D** | |
| | **Estimate** | **Std. Error** | **Estimate** | **Std. Error** |
| $a_{11}$ | 0.9999 | 0.0355 | 0.9927 | 0.0859 |
| $a_{21}$ | -0.0202 | 0.0232 | -0.0135 | 0.0000 |
| $a_{12}$ | 0.2873 | 0.0261 | 0.6779 | 0.0067 |
| $a_{22}$ | 0.8576 | 0.0495 | 0.8523 | 0.0464 |
| $b_{11}$ | 0.0474 | 0.0762 | 0.0615 | 0.0074 |
| $b_{22}$ | 0.0054 | 0.0019 | 0.0053 | 0.0277 |
| $\rho$ | -0.8999 | 0.0133 | -0.8897 | 0.0034 |
| $\overline{\mu}$ | 0.0046 | 0.0011 | 0.0047 | 0.0035 |
| $\overline{\sigma}$ | 0.0440 | 0.0211 | 0.0436 | 0.0002 |
| $\rho_\mu$ | - | - | -0.1902 | 0.0135 |
| $\rho_\sigma$ | -0.9000 | 0.0002 | -0.8942 | 0.0203 |
| $\mathcal{L}$ | 1388.43 | | 1395.5 | |

**Estimation results.** This table describes presents the QML estimates of the model:

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

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

where

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

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.

**Table 2.10.** DSGE Model: Parameter Values

| **Parameter** | $\beta$ | $\delta$ | $\alpha$ | $\rho$ | $\gamma$ | $\sigma_y$ | $\sigma_A$ | $\sigma_c$ |
|---|---|---|---|---|---|---|---|---|
| **Value** | 0.95 | 0.15 | 0.30 | 0.90 | 3.00 | 0.10 | 0.10 | 0.15 |

**Parameter values.** This table contains the parameter values used to analyze the dynamic stochastic general equilibrium model. The parameters are indicated in the first row. The values, used for simulating the series, are given in the second row.

**Table 2.11.** DSGE Model (One Equation): Simulation Results

| | **State-Space Model 1** | | |
|---|---|---|---|
| | MSE($\widehat{k}_t$) | MSE ($\widehat{a}_t$) | MSE ($\widehat{GDP}_t$) |
| **Average** | 0.0855 | 0.0133 | 0.0582 |
| **Std. Dev.** | 0.0284 | 0.0026 | 0.0039 |

**State estimation results.** This table presents the mean and variance of the MSEs between the simulated and the filtered values of the log-investment ($\widehat{k}_t$), the random shocks ($\widehat{a}_t$) and the observable detrended gross domestic product per capita ($\widehat{GDP}_t$). The state-space representation implied by the DSGE model is

$$\begin{aligned}
\widehat{GDP}_t &= \alpha\widehat{k}_t + \widehat{a}_t + v_{y,t},\ v_{y,t} \sim \mathcal{N}\left(0, \sigma_y^2\right), \\
\widehat{k}_{t+1} &= \phi_k\widehat{k}_t + \phi_a\widehat{a}_t + \frac{1}{2}\left(\phi_{kk}\widehat{k}_t^2 + 2\phi_{ak}\widehat{a}_t\widehat{k}_t + \phi_{aa}\widehat{a}_t^2\right) + \frac{1}{2}\phi_{\chi\chi}\chi^2, \\
\widehat{a}_{t+1} &= \rho\widehat{a}_t + \varepsilon_{t+1}.
\end{aligned}$$

The results are based on 500 independent samples of $T = 500$ simulated from the model with parameters shown in Table 2.10 and a perturbation parameter $\chi = 0.10$.

**Table 2.12.** DSGE Model (Two Equations): Simulation Results

| | State-Space Model 2 | | | |
|---|---|---|---|---|
| | MSE($\widehat{k}_t$) | MSE ($\widehat{a}_t$) | MSE ($\widehat{GDP}_t$) | MSE ($\widehat{C}_t$) |
| **Average** | 0.1108 | 0.0185 | 0.0722 | 0.0433 |
| **Std. Dev.** | 0.1471 | 0.0159 | 0.0291 | 0.0423 |

**State estimation results.** This table presents the mean and variance of the MSEs between the simulated and the filtered values of the log-investment ($\widehat{k}_t$), the random shocks ($\widehat{a}_t$), the observables are the detrended gross domestic product per capita ($\widehat{GDP}_t$) and consumption per capita ($\widehat{C}_t$). The state-space representation implied by the DSGE model is

$$\begin{aligned}
\widehat{GDP}_t &= \alpha\widehat{k}_t + \widehat{a}_t + v_{y,t}, \quad v_{y,t} \sim \mathcal{N}\left(0, \sigma_y^2\right), \\
\widetilde{C}_t &= \pi_k\widehat{k}_t + \pi_a\widehat{a}_t + \frac{1}{2}\left(\pi_{kk}\widehat{k}_t^2 + 2\pi_{ak}\widehat{a}_t\widehat{k}_t + \pi_{aa}\widehat{a}_t^2\right) + \frac{1}{2}\pi_{\chi\chi}\chi^2 + v_{c,t}, \quad v_{c,t} \sim \mathcal{N}\left(0, \sigma_c^2\right) \\
\widehat{k}_{t+1} &= \phi_k\widehat{k}_t + \phi_a\widehat{a}_t + \frac{1}{2}\left(\phi_{kk}\widehat{k}_t^2 + 2\phi_{ak}\widehat{a}_t\widehat{k}_t + \phi_{aa}\widehat{a}_t^2\right) + \frac{1}{2}\phi_{\chi\chi}\chi^2, \\
\widehat{a}_{t+1} &= \rho\widehat{a}_t + \varepsilon_{t+1}.
\end{aligned}$$

The results are based on 500 independent samples of $T = 500$ simulated from the model with parameters shown in Table 2.10 and a perturbation parameter $\chi = 0.10$.

**Table 2.13.** DSGE Model: Estimation Results

| Parameter | True Value | Model 1 | | Model 2 | |
|---|---|---|---|---|---|
| | | Average | Std. Dev. | Average | Std. Dev |
| $\beta$ | 0.9500 | 0.8946 | 0.0937 | 0.9835 | 0.0207 |
| $\delta$ | 0.1500 | 0.1965 | 0.0803 | 0.1146 | 0.0307 |
| $\alpha$ | 0.3000 | 0.3990 | 0.2475 | 0.4529 | 0.1203 |
| $\rho$ | 0.9000 | 0.7319 | 0.1207 | 0.6687 | 0.1070 |
| $\gamma$ | 3.0000 | 5.1245 | 3.1691 | 2.3750 | 1.2033 |
| $\sigma_y$ | 0.1000 | 0.0686 | 0.0410 | 0.0904 | 0.0376 |
| $\sigma_A$ | 0.1000 | 0.1018 | 0.0289 | 0.1226 | 0.0298 |
| $\sigma_c$ | 0.1500 | - | - | 0.1631 | 0.0210 |

**Estimation results.** This table describes the sampling distribution of the Quasi-maximum likelihood of the model:

$$\begin{aligned}
\widehat{GDP}_t &= \alpha\widehat{k}_t + \widehat{a}_t + v_{y,t}, \quad v_{y,t} \sim \mathcal{N}\left(0, \sigma_y^2\right), \\
\widetilde{C_t} &= \pi_k\widehat{k}_t + \pi_a\widehat{a}_t + \frac{1}{2}\left(\pi_{kk}\widehat{k}_t^2 + 2\pi_{ak}\widehat{a}_t\widehat{k}_t + \pi_{aa}\widehat{a}_t^2\right) + \frac{1}{2}\pi_{\chi\chi}\chi^2 + v_{c,t}, \quad v_{c,t} \sim \mathcal{N}\left(0, \sigma_c^2\right) \\
\widehat{k}_{t+1} &= \phi_k\widehat{k}_t + \phi_a\widehat{a}_t + \frac{1}{2}\left(\phi_{kk}\widehat{k}_t^2 + 2\phi_{ak}\widehat{a}_t\widehat{k}_t + \phi_{aa}\widehat{a}_t^2\right) + \frac{1}{2}\phi_{\chi\chi}\chi^2, \\
\widehat{a}_{t+1} &= \rho\widehat{a}_t + \varepsilon_{t+1}.
\end{aligned}$$

Model 1 corresponds to the state-space model with the first observation equation only, while Model 2 is the model with two observation equations. The parameters $\beta$, $\delta$, $\alpha$, $\rho$, $\gamma$,$\sigma_y$,$\sigma_A$ and $\sigma_C$ are used as an input to solve a DSGE model with second-order approximations as in Schmitt-Grohe and Uribe (2004). The coefficients $\pi_k$, $\pi_a$, $\pi_{kk}$, $\pi_{ak}$,$\pi_{aa}$,$\pi_{\sigma\sigma}$,$\phi_k$, $\phi_a$, $\phi_{kk}$, $\phi_{ak}$,$\phi_{aa}$ and $\phi_{\chi\chi}$, are the numerical values of these approximations. Finally, the perturbation parameter is $\chi = 0.10$.

**Table 2.14.** Multivariate Stochastic Volatility Model: Parameter Values

| Parameter | $\Phi_{11}$ | $\Phi_{22}$ | $\Sigma_{11}$ | $\Sigma_{21}$ | $\Sigma_{22}$ | $D_{11}$ | $D_{21}$ | $D_{22}$ |
|---|---|---|---|---|---|---|---|---|
| **Value** | 0.9639 | 0.9128 | 0.0415 | 0.0349 | 0.1165 | 0.6077 | -0.5095 | 0.3694 |

**Parameter values.** This table displays the parameter values of the multivariate stochastic volatility model with time-varying variances and correlations. The coefficients are indicated in the first row. The true parameter values, used for simulating the series, are given in the second row.

# Chapter 3

# On the Volatility of the Market Sharpe Ratio

## 3.1 Introduction

The Sharpe ratio measures the excess return of an investment relative to its standard deviation. Most leading consumption-based asset pricing theories imply a relatively stable market Sharpe ratio. However, empirical evidence suggests that there is more variability in the Sharpe ratio than standard models account for. Recently, Lettau and Ludvigson (2010) suggest that the finance literature should address this "Sharpe ratio variability puzzle." They document that the empirical standard deviation of the estimated Sharpe ratio is about 47% per quarter. In contrast, a quarterly calibration of the standard Campbell and Cochrane (1999) model produces a substantially lower volatility of 9%. In turn, Chien, Cole, and Lustig (2012) suggest that passive investors' infrequent rebalancing explains the high variability of market Sharpe ratios.

In this chapter, I examine whether estimates of the variability of the Sharpe ratio might be biased due to limitations of the empirical methodology used in estimation. In particular, I show that measurement error in estimated Sharpe ratios may help to explain the Sharpe ratio volatility puzzle. To do this, I simulate data from a standard calibration of the Bansal and Yaron (2004) long-run risks (LRR) model. Following practice common in the literature, I then estimate Sharpe ratios using ordinary least squares (OLS) methods to infer the variability of the model-generated Sharpe ratios. OLS methods lead to estimates of Sharpe ratio volatility of approximately 18%, even though the true variability of the model-implied Sharpe ratio is only 3%. The difference in estimates is due to measurement error induced by the standard Sharpe ratio estimation methodology.

Once I have documented the difference between the Sharpe ratio's estimated and true volatility, I study whether improved empirical methodologies might better account for the true

variability of the Sharpe ratio. In particular, I implement filtering methods, which are statistical tools that recover unobservable state variables using measurements that are observed with noise. These techniques are flexible enough to allow the econometrician to perform statistical inference based on time-varying information observed with measurement error. Moreover, the modeling representation is general enough to include time varying information as well as flexible correlation structures and errors in variables.[38]

I use two different exercises to show the limitations of OLS methods. First, I run a controlled experiment in which I have full information of the data–generating process of stock returns, state variable dynamics and parameter values. I use simulated data from the LRR model as calibrated by Bansal and Yaron (2004) to estimate conditional means, variances and Sharpe ratios. The use of artificial data from a fully specified economy is important because it allows the economic reasons that drive the variation in Sharpe ratios to be isolated. Moreover, information about model specification and state-variable dynamics is incorporated within the filtering estimation procedure. Furthermore, the tractability of the LRR model allows data to be simulated with relative ease.

I then implement two econometric techniques: I run standard OLS regressions and then I apply filtering techniques. I compare both sets of results with the closed-form expressions implied by the LRR model as a benchmark. My results show that the Sharpe ratios based on standard OLS methods are more volatile than the estimates obtained with filtering techniques. Moreover, the volatility estimates obtained via filtering differ from the true value by less than 1%, which is a significant improvement over OLS estimates. The main driver of this result is the use of conditioning information within the estimation process.

There are a number of reasons why a filtering approach can improve upon predictive regressions to estimate expected returns and conditional volatilities. First, filtering explicitly acknowledges that both expected returns and volatilities are time varying. Filtering techniques aggregate the entire history of realized returns parsimoniously; in contrast, predictive regressions use lagged predictors to form estimates of expected returns and volatilities. Instead of adding lags to a vector autorregressive (VAR) model, which would increase the number of parameters to be estimated, a latent variable approach such as filtering incorporates the information contained in the history of observed returns. Moreover, filtering techniques are flexible enough to be used with large information sets without relying on additional instruments that may be misspecified (Ferson, Sarkissian, and Simin, 2003). Finally, filtering is more robust to structural breaks than are OLS techniques (Rytchkov, 2012), since it is insensitive to robust shifts in relations over the long run. For example, in the predictability literature, a substantial shift in the dividend-price ratio destroys its forecasting power.[39] Also, robustness to structural breaks makes the filtering

[38]See Hamilton (1994); Kim and Nelson (1999) and Doucet, de Freitas, Gordon, and Smith (2001) for an introduction to filtering methods. Crisan and Rozovskii (2011) provide a more recent literature review in nonlinear filtering methods.

[39]See Lettau and Van Nieuwerburgh (2008) for a detailed explanation.

approach more valuable from an ex-ante point of view, when it is unclear whether structural breaks will occur.

The standard method used in the literature to estimate Sharpe ratios is to use fitted moments from first-stage predictive regressions as proxies for the unobserved conditional mean and volatility. Such a technique has some important drawbacks. First, the dynamics of the conditional mean and volatility are determined by the joint conditional distribution of the first-stage predictors. Thus, with any model misspecification, such as omitted variables, the dynamics of the fitted moments would not necessarily correspond to the dynamics of the true moments. In addition, even if the predictive models for the conditional mean and volatility are well specified, the effect of errors in variables, which are induced by the first-stage regressions, is not trivial to quantify in a VAR model.

Simulating data of stock returns by means of theoretical models is a powerful tool because the economic reasons that drive the simulated time-series variation are fully identified. However, theoretical models are abstractions, and by definition misspecified. An alternative form of analyzing stock returns is via reduced form models, which are statistical representations that do not impose any economic structure and thus aim to better describe historical data. To infer Sharpe ratios and their variability from the data, I conduct a second exercise based on the reduced form model by Brandt and Kang (2004). In this model, expected returns and volatilities are estimated as latent variables and identified from the history of returns. The main advantage of this approach is that it does not rely on prespecified predictors and is not subject to errors in variables or model misspecification. I apply the filtering techniques described in Chapter 2 to estimate the parameters of the model and to extract estimates of conditional moments of returns as well as conditional Sharpe ratios. As a result, my estimate for quarterly Sharpe ratio volatility using the reduced form model is in the order of 5% to 10%, whereas my estimate for the quarterly Sharpe ratio volatility using the OLS methods is 42%.

Consistent with the results of the simulation exercise, I find that conditioning information drives the results above. Reduced form models do not rely on predetermined conditioning variables to estimate conditional moments: The state variables are identified from the history of returns. Standard OLS techniques generate fitted moments from a set of predictive regressions as proxies for the unobservable conditional mean and volatility. The fitted moments depend on the joint distribution of these predictors. Consequently, any model misspecification would generate fitted moments that do not correspond to the true dynamics of the conditional mean and volatility, and thus, the dynamics of the Sharpe ratio.

My findings have important implications in an asset management context since the Sharpe ratio is a commonly used measure of performance evaluation. For investors willing to allocate their wealth between the market portfolio and the risk-free instrument, the market Sharpe ratio becomes a natural benchmark of their investments. If this ratio is highly volatile, the

variation needs to be taken into account for hedging and rebalancing purposes. Indeed, Lustig and Verdelhan (2012) report that accounting for time variation in Sharpe ratios may lead to optimal trading strategies that differ markedly from buy-and-hold strategies.

Furthermore, a mean-variance investor would have an obvious interest in understanding the volatility of Sharpe ratios. For example, in a partial equilibrium setting,[40] the Sharpe ratio determines the fraction of wealth that an agent invests in the market portfolio. I show that if an investor uses OLS methods to determine this fraction, then the portfolio weights would exhibit extremely volatile behavior over time, which may result in high rebalancing costs. I also show that if an investor applies filtering techniques to estimate the fraction of wealth invested in the market portfolio, then these costs will be substantially lower. Even further, for a representative agent with habit formation preferences, the Sharpe ratio indicates the timing and magnitude of fluctuations of risk aversion (Campbell and Cochrane, 1999; Constantinides, 1990). Thus, the time variation in the market Sharpe ratio may provide information about the fundamental economics underlying the asset prices.

### 3.1.1 Related Literature

A number of studies analyze the predictable variation of the mean and volatility of stock returns from an empirical point of view.[41] However, only a few papers have investigated the time variation observed in equity Sharpe ratios. Lettau and Ludvigson (2010) measure the conditional Sharpe ratio of U.S. equities by forecasting stock market returns and realized volatility using different predictors. They obtain highly counter-cyclical and volatile Sharpe ratios and show that neither the external habit model of Campbell and Cochrane (1999) nor the LRR model Bansal and Yaron (2004) deliver Sharpe ratios volatile enough to match the data. Using a latent VAR process, Brandt and Kang (2004) also find a highly counter-cyclical Sharpe ratio. Ludvigson and Ng (2007) document the same result using a large number of predictors in a dynamic factor analysis.

Tang and Whitelaw (2011) document predictable variation in stock market Sharpe ratios. Based on a predetermined set of financial variables, the conditional mean and volatility of equity returns are constructed and combined to estimate the conditional Sharpe ratios. Tang and Whitelaw (2011) find that conditional Sharpe ratios show substantial time variation that coincides with the phases of the business cycle. Lustig and Verdelhan (2012) provide evidence that Sharpe ratios are higher in recessions than in expansions in the United States and other OECD countries. They also find that the changes in expected returns during business-cycle expansion and contractions are not explained by changes in near-term dividend growth rates. These papers focus on the counter-cyclical behavior of Sharpe ratios. My paper focuses on

[40] Some examples are Merton (1969, 1971).

[41] See Lettau and Ludvigson (2010) for a comprehensive survey.

the conditional volatility of market Sharpe ratios and finds that the volatility estimates are substantially smaller than the evidence previously documented.

My paper is also related to Brandt and Kang (2004); Pástor and Stambaugh (2009); van Binsbergen and Koijen (2010) and Rytchkov (2012), who analyze return predictability using state-space models.[42] I contribute to the literature by focusing on the dynamic behavior of the market Sharpe ratio and by showing that standard OLS methods as applied in the literature generate measurement error which impacts estimates of Sharpe ratio volatility. Moreover, I also show that filtering techniques are a good approach for estimating the ratio's true volatility. I also find that filtering techniques are better able to capture the dynamic behavior of market Sharpe ratios.

The remainder of this chapter is organized as follows. Section 3.2 provides a theoretical framework to interpret Sharpe ratios. Section 3.3 introduces the LRR model and its implications for empirical moments. Section 3.4 describes the simulation exercise as well as the estimation methodologies for expected returns, volatilities and Sharpe ratios. In section 3.5, an analysis of historical Sharpe ratios based on reduced form models is described and the empirical results are shown. Section 3.6 presents asset allocation implications. Finally, concluding remarks are presented in section 3.7.

## 3.2 Sharpe Ratios in Asset Pricing

The conditional Sharpe ratio of any asset at time $t$, denoted by $SR_t$, is defined as the ratio of the conditional mean excess return to its conditional standard deviation; that is,

$$SR_t = \frac{\mathbb{E}_t\left[R_{t+1} - R_{ft+1}\right]}{\sigma_t\left[R_{t+1} - R_{ft+1}\right]}, \tag{3.1}$$

where $R_t$ and $R_{ft}$ denote the gross asset return of an asset and the one-period risk-free interest rate, respectively, and the conditional expectations are based on the information available at time $t$.

Harrison and Kreps (1979) show that the absence of arbitrage implies the existence of a stochastic discount factor (SDF) or pricing kernel, denoted by $M_t$, that prices all assets in the economy.[43] An implication of no arbitrage is that the expectation of the product of the stochastic discount factor and the gross asset return of any asset must be equal to one; that is,

$$\mathbb{E}_t\left[M_{t+1}R_{t+1}\right] = 1. \tag{3.2}$$

[42] In an early work in this body of literature, Conrad and Kaul (1988) use the Kalman filter to extract expected returns, but only from the history of realized returns. Other studies that relate latent variables with predictability include Ang and Piazzesi (2003); Lamoureux and Zhou (1996) and Dangl and Halling (2012).

[43] A detailed explanation is shown in Appendix B.1.

An implication of (3.2) is that the conditional Sharpe ratio is proportional to the risk-free rate, the volatility of the pricing kernel and the correlation between the pricing kernel and the return; that is,

$$SR_t = -R_{ft+1}\sigma_t\left[M_{t+1}\right] Corr_t\left[R_{t+1}, M_{t+1}\right], \tag{3.3}$$

where $\sigma_t$ and $Corr_t$ are the standard deviation and correlation, conditional on information at time $t$, respectively. The conditional Sharpe ratio of any asset in the economy is time varying as long as the risk-free rate varies or the pricing kernel is conditionally heteroskedastic -that is, $\sigma_t\left[M_{t+1}\right]$ changes over time- or if the correlation between the stock market return and the SDF is time varying. In this paper, I focus on the conditional Sharpe ratio of the aggregate stock market, which is defined as the instrument that pays the aggregate dividend every period. However, the analysis can be extended to the Sharpe ratios of any traded asset.

The next section presents the LRR model of Bansal and Yaron (2004), with a particular focus on the implications for expected returns, volatilities and Sharpe ratios of the aggregate stock market. This model explains stock price variation as a response to persistent fluctuations in the mean and volatility of aggregate consumption growth by a representative agent with a high elasticity of intertemporal substitution. The tractability of the LRR model allows data to be simulated with relative ease. It provides analytical expressions for expected returns, volatilities and Sharpe ratios for the market portfolio, conditional on the Campbell and Shiller (1988) log-linearizations. Later in the paper I briefly present other asset pricing models and their implications for market Sharpe ratios.

## 3.3 The Long-Run Risks Model

Bansal and Yaron (2004) and Bansal, Kiku, and Yaron (2012a) (BY and BKY hereafter) propose the following stochastic processes for the log-consumption and log-dividend growth, denoted by $\Delta c_{t+1}$ and $\Delta d_{t+1}$, respectively:

$$\begin{aligned}
\Delta c_{t+1} &= \mu_c + x_t + \sigma_t\eta_{t+1} \\
x_{t+1} &= \rho x_t + \varphi_e\sigma_t e_{t+1} \\
\sigma^2_{t+1} &= \overline{\sigma}^2 + v\left(\sigma^2_t - \overline{\sigma}^2\right) + \sigma_w w_{t+1} \\
\Delta d_{t+1} &= \mu_d + \phi x_t + \varphi\sigma_t u_{t+1} + \pi\sigma_t\eta_{t+1} \\
w_{t+1}, e_{t+1}, u_{t+1}, \eta_{t+1} &\sim i.i.d.\ \mathcal{N}(0,1),
\end{aligned} \tag{3.4}$$

where, $x_t$ is a persistently varying component of the expected consumption growth rate and $\sigma^2_t$ is the conditional variance of consumption growth, which is time varying and highly persistent, with unconditional mean $\overline{\sigma}^2$. The variance process can take negative values, but it will happen with small probability if its conditional mean is high enough with respect to its variance. Dividends are correlated with consumption since the growth rate, $\Delta d_{t+1}$, shares the same

persistent predictable component scaled by a parameter $\phi$, and the conditional volatility of dividend growth is proportional to the conditional volatility of consumption growth.

BY solve the LRR model using analytical approximations. They assume a representative agent with Epstein-Zin utility with time discount factor $\delta$, coefficient of relative risk aversion $\gamma$, and elasticity of intertemporal substitution $\psi$. The log of the stochastic discount factor, $m_{t+1}$, for this economy is given by

$$m_{t+1} = \theta \ln \delta - \frac{\theta}{\psi} \Delta c_{t+1} + (\theta - 1) r_{a,t+1}, \tag{3.5}$$

where $\theta = (1-\gamma)/(1-\psi)$ and $r_{a,t+1}$ is the return on the consumption claim, or equivalently, the return on aggregate wealth. BY use the Campbell and Shiller (1988) log-linearizations to obtain analytical approximations for the returns on the consumption and dividend claims. Further details on the model and derivations are explained in Appendix B.2.

### 3.3.1 Implications for Expected Returns, Volatilities and Conditional Sharpe Ratios

Under the long-run risks framework, the equity premium is an affine function of the volatility of consumption growth alone:

$$\mathbb{E}_t \left[ r_{m,t+1} - r_{f,t+1} \right] = E_0 + E_1 \sigma_t^2. \tag{3.6}$$

The model also implies that the conditional variance of the market return is an affine function of $\sigma_t^2$ :

$$Var_t \left( r_{m,t+1} \right) = D_0 + D_1 \sigma_t^2. \tag{3.7}$$

The coefficients $E_0, E_1, D_0$ and $D_1$ are known functions of the underlying time-series and preference parameters. The general expressions and details about their derivation are shown in Appendix B.2.4.

The covariance between the observed market excess return, $r_{m,t+1}$, and the innovation to the volatility process, $w_{t+1}$ is given by

$$cov_t \left( r_{m,t+1}, w_{t+1} \right) = \kappa_{1,m} A_{2,m} \sigma_w. \tag{3.8}$$

One of the appealing properties of the long-run risk model, is that $A_{2,m} < 0$, for standard calibrations, implying that the LRR model is able to reproduce the negative feedback effect.[44] Another implication of Eq. (3.8) , is that the conditional correlation between excess returns and

[44] Campbell and Hentschel (1992b), Glosten, Jagannathan, and Runkle (1993), and Brandt and Kang (2004), among others document the volatility feedback effect; that is, return innovations are negatively correlated with innovations in market volatility.

the innovations to consumption risk is time varying because the conditional variance of stock returns is time varying.

Following Bansal, Kiku, and Yaron (2012b), the cumulative log-return over $K$ time periods is just a sum of $K$ one-period returns,[45]

$$\sum_{k=1}^{K} \left( r_{m,t+k} - r_{f,t+k} \right).$$

The conditional moments are given by

$$\mathbb{E}_t \left[ \sum_{k=1}^{K} r_{m,t+k} - r_{f,t+k} \right] = \mathbf{E}_{0,K} + \mathbf{E}_{1,K} \sigma_t^2, \tag{3.9}$$

and

$$Var_t \left[ \sum_{k=1}^{K} r_{m,t+k} - r_{f,t+k} \right] = \mathbf{D}_{0,K} + \mathbf{D}_{1,K} \sigma_t^2, \tag{3.10}$$

where $\mathbf{E}_{0,K}$, $\mathbf{E}_{1,K}$, $\mathbf{D}_{0,K}$, and $\mathbf{D}_{1,K}$ are known functions of the preference parameters and the number of periods, $K$, used for time aggregation. If the time unit is the month, then evaluating Eqs. (3.9) and (3.10) provides an expression for annual estimates.

The conditional Sharpe ratio of an investment over $K$ time periods is given by the ratio of the conditional mean returns divided by its conditional standard deviation, and is represented by

$$SR_{t,t+K} = \frac{\mathbf{E}_{0,K} + \frac{\mathbf{D}_{0,K}}{2} + \left( \mathbf{E}_{1,K} + \frac{\mathbf{D}_{1,K}}{2} \right) \sigma_t^2}{\sqrt{\mathbf{D}_{0,K} + \mathbf{D}_{1,K} \sigma_t^2}}. \tag{3.11}$$

Eq. (3.11) implies that the only source of variation in the conditional Sharpe ratio under the LRR framework is the volatility of consumption growth. Moreover, the conditional Sharpe ratio is stochastic unless $\sigma_t^2$ is deterministic. Furthermore, under the standard calibrations by BY, the conditional Sharpe ratio is strictly increasing in the volatility of consumption growth. This implies that the long-run risk framework predicts counter-cyclical Sharpe ratios; that is, for bad times (high values of the volatility of consumption growth) the Sharpe ratios are high and for good times, (low values of volatility of consumption growth) the conditional Sharpe ratios are low, consistent with the habit formation model of Campbell and Cochrane (1999).

Moreover, Eqs. (3.9) , (3.10) and (3.11) characterize the expected return, volatility and Sharpe ratios of a buy and hold strategy over $K$ time units. These equations define the term

[45]Time aggregation is an important mechanism for parameter and state inference. Bansal, Kiku, and Yaron (2012b) explicitly consider time aggregation of variables. They find that time aggregation can affect parameter values and they provide evidence that ignoring time aggregation leads to false rejection of the LRR model. Earlier papers that account for time aggregation in estimation in asset pricing context include Hansen and Sargent (1983) and Heaton (1995).

structure of risk premia, volatility, and Sharpe ratios of the market portfolio. Moreover, by evaluating Eqs. (3.9) and (3.10) in the unconditional value of the volatility of consumption growth, $\sigma^2$, we obtain expressions for the unconditional moments of cumulative returns. Similar expressions can be obtained for the cumulative return moments of the risk-free instrument and market portfolio. Details about the derivations are described in Appendix B.3.

## 3.4 Sharpe Ratios Simulated from Structural Models

In this section, I conduct a simulation study in the spirit of Beeler and Campbell (2012). The objective is to simulate equity returns from the LRR model at a monthly frequency, and then time aggregate them to obtain annual estimates of returns, volatilities and Sharpe ratios. I explain the simulation exercise as follows.

First, I generate four sets of independent standard normal random variables and use them to construct monthly series for consumption, dividends and state variables using the state-space model in Eq. (3.4).[46] Next, I construct annual consumption and dividend growth by adding twelve monthly consumption and dividend levels, respectively, and then take the growth rate of the sum. The log market returns and risk-free rates are the sum of monthly values, while the log price-dividend ratios use prices measured from the last period of the year. As the price-dividend ratio in the data is divided by the previous year's dividends, the price-dividend ratio in the model is multiplied by the dividend in that month and divided by the dividends over the previous year.

As in BY, BKY and Beeler and Campbell (2012), negative realizations of the conditional variance are censored and replaced with a small positive number.[47] I also retain sample paths along which the volatility process goes negative and is censored.[48] Since the volatility is highly persistent, it is quite likely to have negative values for the conditional variance; indeed, Beeler and Campbell (2012) report that under the BK calibration less than 1% of the volatility simulations are negative for a sample of 100,000 simulations. Each simulation is initialized from the steady-state values and run for a "burn-in" period of ten years.

### 3.4.1 Predictive Regressions

The conditional moments of market returns as well as the Sharpe ratio are unobservable. A common approach that has been applied in the empirical literature to circumvent this issue is to project excess stock returns series on a predetermined set of conditioning variables, such as economic or financial indicators observed by the econometrician.

[46] The frequency is consistent with the parameters calibrated by BY and BKY, which are provided in monthly terms.

[47] The number is $(10^{-14})$ and is consistent with the simulation exercise of Beeler and Campbell (2012).

[48] An alternative approach is to replace negative realizations with their absolute values, as in Johnson and Lee (2012).

Empirical studies differ in the conditioning information used in projection of excess returns. Among the most commonly used predictor variables are the price-dividend ratios (Campbell, 1991; Fama and French, 1988a; Hodrick, 1992), short-term interest rates (Ang and Bekaert, 2007; Campbell, 1991; Fama and Schwert, 1977; Hodrick, 1992), term spreads and default spreads (Fama and French, 1988a), book market ratios (Lewellen, 1999; Vuolteenaho, 2000), proxies for consumption-wealth ratio (Lettau and Ludvigson, 2001a,b), and latent factors obtained from large data sets (Ludvigson and Ng, 2007).[49] Expected returns are calculated by regressing realized returns on the set of predictors and taking the fitted values as estimates.

Conditional volatility may also be measured by a projection onto predetermined conditioning variables, taking the fitted value from this projection as a measure of conditional variance or conditional standard deviation. This type of modeling is commonly used; for example French, Schwert, and Stambaugh (1987) use a time-series model of realized variance to model the conditional variance.

Within the set of techniques to measure conditional volatility by a projection onto predetermined conditioning variables, three approaches are common. One is to take the squared residuals from a regression of excess returns onto a predetermined set of conditioning variables and regress them on to the same set of conditioning variables, using the fitted values from this regression as a measure of conditional variance.[50] Alternatively, volatility can be estimated using high-frequency return data, commonly referred to as realized volatility. This is an ex-post measure that consists of adding up the squared high-frequency returns over the period of interest. The realized volatility is then projected onto time $t$ information variables to obtain a consistent estimate of the conditional variance of returns.[51] The third approach estimates conditional volatility of excess stock market returns by specifying a parametric form for the conditional volatility, such as the GARCH type of models, or stochastic volatility.[52] The volatility estimates are then obtained from the history of observed returns. For this part of the paper, I focus on the second type of methodology to calculate conditional volatilities of stock returns by projecting the sum of squared monthly returns on a set of predictors.

As for the conditional Sharpe ratio, a standard measure is the ratio of the estimated expected excess return to the estimated volatility, both obtained from separate projections. This approach has been taken by Kandel and Stambaugh (1990), Tang and Whitelaw (2011) and Lettau and Ludvigson (2010), among others.

[49] Goyal and Welch (2008); Lettau and Ludvigson (2010) provide a comprehensive review of predictive variables commonly used in the literature.

[50] Campbell (1987) and Breen, Glosten, and Jagannathan (1989) apply these methods in the predictability literature.

[51] This approach is taken by French, Schwert, and Stambaugh (1987), Schwert (1989), Whitelaw (1994), Ghysels, Santa-Clara, and Valkanov (2006), Ludvigson and Ng (2007), Lettau and Ludvigson (2010) and Tang and Whitelaw (2011).

[52] French, Schwert, and Stambaugh (1987), Bollerslev, Engle, and Wooldridge (1988) and Glosten, Jagannathan, and Runkle (1993) have applied this approach in the predictability literature.

I model the conditional moments of annual returns as follows:

$$\mathbb{E}_t\left[R_{t+1} - R_{f,t+1}\right] = X_t\beta_\mu, \tag{3.12}$$

$$Var_t\left[R_{t+1} - R_{f,t+1}\right] = X_t\beta_\lambda, \tag{3.13}$$

where $X_t$ is the set of predictor variables observed at time $t$ and $R_{t+1} - R_{f,t+1}$ is the annual excess return on the market. I assume that the predictor variables available at time $t$ are the price-to-dividend ratio, the current excess returns and the risk-free rate, constructed in an annual basis.

The regression equations that correspond to (3.12) and (3.13) are

$$R_{t+1} - R_{f,t+1} = X_t\beta_\mu + \varepsilon_{\mu,t+1}, \tag{3.14}$$

$$v_{t+1} = X_t\beta_\lambda + \varepsilon_{\lambda,t+1}, \tag{3.15}$$

where $R_t - R_{f,t}$ is the annual excess return on the market portfolio and $v_t$ is the realized variance for year $t$. The annual excess return is calculated as the sum of the monthly excess log-returns, while the realized variance is the sum of the squared monthly excess log-returns. Both sums are calculated within the same year.

Based on the information available at time $t$ and the parameter estimates from (3.12) and (3.13), the conditional Sharpe ratio is calculated as follows:

$$\widehat{SR}_t = \frac{X_t\widehat{\beta}_\mu + \frac{X_t\widehat{\beta}_\lambda}{2}}{\sqrt{X_t\widehat{\beta}_\lambda}}, \tag{3.16}$$

where $\widehat{\beta}_\mu$ and $\widehat{\beta}_\lambda$ denote the OLS estimates for $\beta_\mu$ and $\beta_\lambda$ respectively.[53]

Figure 3.1 shows the results of a simulated path of annual returns. Each simulation has 100 annual observations of returns. Panel A shows the time series of expected returns calculated from an OLS regression, Panel B shows the conditional variance estimated from an OLS regression and Panel C contains the conditional Sharpe ratio estimates using the fitted values from the conditional mean and conditional volatility from panels A and B. Finally, Panel D displays the time series of annual Sharpe ratios implied by the BY model. These are obtained by evaluating Eq. (3.11) in $K = 12$.

For this specific simulation, the standard deviation of the Sharpe ratio estimates is 3% while the standard deviation of the model Sharpe ratio is 17%. Moreover, the correlation coefficient between the Sharpe ratio estimates based on OLS methods and the Sharpe ratio implied by the model is 7.9%.

[53]This definition of Sharpe ratio includes the Jensen's adjustment due to log-returns. However, my results are robust if the Sharpe ratio is defined as the ratio of expected returns to conditional volatility.

The use of artificial data from a fully specified economy is important because it allows the economic reasons that drive the variation in Sharpe ratios to be isolated. In the first case, the variation in the Sharpe ratio is driven by the volatility of consumption growth. In the second case, the volatility of the Sharpe ratio is driven by consumption risk and measurement error caused by the OLS estimation method. In order to verify the robustness of my results, I repeated the previous exercise 100,000 times via Monte Carlo simulations with sample periods of length 100 years. Table 3.2 reports the median moments implied by the simulations of the BY calibrations of the LRR calibrations. I look at the empirical first and second moments and at the empirical Sharpe ratios constructed via OLS methods and compare them with the median first and second moments as well as the Sharpe ratios implied by the model.

From Table 3.2, we learn that the level of expected returns and volatilities implied by the LRR model are well captured by OLS techniques. Indeed, the difference between expected returns and the LRR model counterpart is almost indistinguishable. As for the volatility estimates, OLS techniques do a good job in matching the mean level as well as standard deviation. However, there are some differences worth noting. The standard deviation of the risk premia calculated with OLS techniques is 3.01%, while the standard deviation implied by the model is 0.87%. That is, the standard deviation estimated via OLS methods is more than three times the true standard deviation. Moreover, the median of the correlation coefficient between the risk premia and its OLS estimate is 0.52%. A more serious discrepancy is observed in the estimates of the conditional Sharpe ratio. The model implies a median annual Sharpe ratio of 33.33% while the estimates obtained with projection techniques is 26.45%; the standard deviation of the Sharpe ratio calculated regressions is 15.82%, while, the value implied by the model is 3.53%. The correlation between the true Sharpe ratios implied by the model and its OLS estimates is 0.39%.

We learn from this simulation exercise that the use of fitted moments as proxies for the unobserved conditional mean and volatility of stock returns has some obvious drawbacks. First, the dynamics of the conditional mean and volatility are determined by the joint conditional distribution of the first-stage predictors. Thus, with any model misspecification the dynamics of the fitted moments would not need to correspond to the dynamics of the true moments. Even when the predictive models for the conditional mean and volatility are well specified, the effect of errors in variables, which are induced by the first stage-regressions, is not trivial to quantify and has an important effect in the Sharpe ratio volatility estimates. Moreover, OLS methods do not account for time-varying observations or time-varying information sets; therefore OLS methods are not robust to structural changes. In that sense, an econometric technique that accounts for such deficiencies may be a good approach for Sharpe ratio estimation and its dynamic behavior. Filtering techniques are able to overcome these issues.

### 3.4.2 Filtering and Estimation

Let $y_{t+1} = r_{m,t+1} - r_{f,t+1}$ be the continuously compounded monthly excess return. The time-series dynamics of $y_{t+1}$ is represented by

$$y_{t+1} = \mu_t + \lambda_t \varepsilon_{t+1} \text{ with } \varepsilon_{t+1} \sim \mathcal{N}(0,1), \tag{3.17}$$

where $\mu_t$ and $\lambda_t$ represent the expected return and conditional volatility. Under the LRR model, these are given by

$$\mu_t = E_0 + E_1 \sigma_t^2, \tag{3.18}$$

and

$$\lambda_t^2 = D_0 + D_1 \sigma_t^2. \tag{3.19}$$

According to Eq. (3.4), the evolution of $\sigma_t^2$ is represented by

$$\begin{aligned} \sigma_{t+1}^2 &= \overline{\sigma}^2 + v\left(\sigma_t^2 - \overline{\sigma}^2\right) + \sigma_w w_{t+1}, \\ w_{t+1} &\sim i.i.d.\ \mathcal{N}(0,1), \end{aligned} \tag{3.20}$$

and the covariance between the observed market excess return, $y_{t+1}$, and the innovation to the volatility process, $w_{t+1}$ is given in Eq. (3.8).

#### Filtering

Eqs. (3.17) through (3.20) make up a state-space model. In the terminology of state-space models, Eq. (3.17) is the measurement or observation equation and Eq. (3.20) is the transition or state equation. I assume that $\sigma_t^2$ is a latent variable; therefore, both the conditional mean and volatility of market returns are unobservable. I also assume that I am able to observe the full history of realized returns. To draw inferences about the dynamic behavior of $\sigma_t^2$ as well as the conditional distribution of excess returns, we need to solve a filtering problem.

The solution to the filtering is the distribution of the latent variable $\sigma_t^2$ conditional on the history of observed returns. From Eqs. (3.9) through (3.11), we learn that expected returns, volatilities and conditional Sharpe ratios can be estimated based on this conditional distribution, for any holding period. Unfortunately, the filtering problem generated by the LRR model is not standard because of the nonlinearities in the measurement equation as well as the non-zero covariance between the observation and transition equations. As a result, the standard Kalman filter (designed for linear Gaussian state-space models) cannot be used directly in the estimation of the model. I instead rely on nonlinear filtering methods to estimate the distribution of $\sigma_t^2$, conditional moments of market excess returns and market Sharpe ratios.

**Particle Filters**

I estimate the latent process $\sigma_t^2$, conditional moments and Sharpe ratios via particle filters. The particle filter is a nonlinear filter which works through Monte Carlo methods. The conditional distribution of the state variables is replaced by an empirical distribution drawn by simulation. This method does not require the explicit computation of Jacobians and Hessians, and captures the conditional distribution of the state variable accurately up to a prespecified accuracy level that depends on the number of simulations chosen by the researcher. To implement the particle filter, it is necessary to specify the state-space model.[54] A brief description of the particle filter and its implementation is given in Appendix A.6.

I test the accuracy of the filtered estimates as follows. First, I simulate a path of annual excess returns according to the calibrations by BY. Given the simulated excess returns, I numerically construct the conditional distribution of the volatility of consumption growth, $\sigma_t^2$, using Eqs. (3.17) to (3.20) as well as the original calibrations by BY. Once, the conditional distribution of the volatility of consumption growth is obtained, I estimate risk premia, conditional variances and Sharpe ratios according to Eqs. (3.9) to (3.11). Figure 3.2 shows a sample simulation of the volatility of consumption growth, conditional moments and Sharpe ratios along with their filtered counterparts. In panel A, I show a path for the volatility of consumption growth; panels B and C show the simulated expected returns and their volatility; panel D shows the simulated annual Sharpe ratio with its filtered estimates. For this specific simulation, the correlation coefficient between the simulated volatility of consumption growth and its filtered value is 60%. As for the expected returns, volatilities and Sharpe ratios, the simulated values have a correlation coefficient above 64% with their filtered counterparts.

A common concern is that filtering is thought of as a smoothing technique, and therefore, if the state variable to be filtered is too volatile, a filtering technique will reduce such volatility and the unconditional moments of interest may not reflect the true state variable dynamics. However, filtering techniques are robust enough to provide accurate estimates even if the true state variable to be filtered is volatile. This is due to the fact that filtered estimates are conditional expectations of the state variables. To evaluate the unconditional moments, it is necessary to account for this fact; thus, I calculate the unconditional mean and variance of the state variables according to the properties of the law of iterated expectations.[55]

To verify the robustness of my results, the simulation exercise was repeated 1,500 times. For each simulation, I obtain time series of expected returns, volatilities and Sharpe ratios as well as their filtered counterparts. I then calculate the unconditional means, variances and correlations

[54] Doucet, de Freitas, Gordon, and Smith (2001) and Crisan and Rozovskii (2011) describe in detail the properties of the filter and its practical implementation, and van Binsbergen, Fernandez-Villaverde, Koijen, and Rubio-Ramirez (2012) apply the method to estimate a dynamic stochastic general equilibrium model with a particular focus on the term structure of interest rates.

[55] The unconditional variance estimate comes from the following identity, which relates conditional and unconditional variances: $Var[X] = Var[\mathbb{E}[X|Y]] + \mathbb{E}[Var[X|Y]]$.

between the simulated and filtered series. The results are reported in Table 3.3. In general, the filters do a good job of capturing the unconditional moments of expected returns, volatilities and Sharpe ratios. Overall, the moments estimated via filtering methods are precise match the true values in at least two decimal places.

We learn from these simulation exercises that filtering techniques are better able to capture the dynamic behavior of the conditional moments and Sharpe ratios than OLS methods. Nonetheless, these filtered estimates rely on a number of assumptions: the state-space model is well specified; realized returns are a noisy measure of expected returns and the volatility of consumption growth is the only unobservable state variable of the system. However, the researcher has full knowledge of its dynamics as well as the functional forms of expected returns and variances. Finally, I assume that the econometrician has full knowledge of the parameter values and the only problem that she faces is the estimation of conditional moments based on the time series of observed returns. By using OLS methods, we approximate expected returns and variances with a linear projection on a set of exogenous predictors and can potentially face a number of well-known econometric problems, such as omitted variables or misspecification. For clarity of exposition, I collect all parameters that define the state-space model into a single parameter vector $\theta$. Each parameter vector characterizes a model; hence, conditional distributions and filtered state variables. As a result, an estimation problem needs to be solved and will be explained in detail as follows.

**Estimation**

The previous results were obtained by assuming that the set of parameter values is known. This assumption is quite unrealistic, because in reality the researcher is uncertain about the true parameter values. A natural way to approach this issue is by estimating the vector of parameters from the observed data. A common technique for nonlinear dynamic models is QML estimation,[56] as described in Chapter 2. Details about its implementation are described in Appendix B.4. I conduct a simulation exercise to better identify the effect of parameter estimation within the filtering exercise. First, I simulate a time series of excess returns from the LRR model, and then I estimate the parameter values via QML estimation methods using the state-space representation implied by the LRR model. The parameter estimates are then used in the filtering estimation procedure.

A sample simulation is illustrated in Figure 3.3. Panels 3.3a to 3.3c compare conditional Sharpe ratio estimates with their true values. Panel 3.3a shows the empirical estimate obtained via OLS methods, Panel 3.3b displays Sharpe ratio estimates calculated with filtering methods by assuming that the true parameter values are known and Panel 3.3c contains the filtered Sharpe ratios using the parameter estimates obtained via QML methods.

[56]Some examples are Campbell, Sunderam, and Viceira (2012); van Binsbergen and Koijen (2011) and Calvet, Fisher, and Wu (2013).

For this specific simulation, the time-series average of the model-implied Sharpe ratio is 33%, while the average Sharpe ratio estimates obtained with filtering techniques are 34% and 36%, where the first is obtained by assuming that the true parameter values are known and the second is obtained with the parameter estimates from observed returns. Finally, the average Sharpe ratio obtained with OLS methods is 25%. An explanation for this difference is the model misspecification that is generated from running OLS regressions for expected returns and volatility calculations on a set of predetermined variables. The volatility estimates obtained via filtering methods are 5% and 6%. My results are similar to the estimates obtained from the true simulated data, which is 4%. In contrast, OLS methods deliver a Sharpe ratio volatility estimate of 15%. This exercise illustrates the effect of parameter estimation on the volatility of Sharpe ratios. I show evidence that filtering methods deliver Sharpe ratio volatility estimates consistent with the true model implied values, even if parameter values have to be estimated.

### 3.4.3 Other Models

Recent consumption-based asset pricing models have made substantial progress in explaining many asset pricing puzzles across various markets. Even though such models are not often used to study Sharpe ratios or their volatility, they do make theoretical predictions about their values. In standard asset pricing models, the market Sharpe ratio is constant (Breeden, 1979; Lintner, 1965; Lucas, 1978; Sharpe, 1964) or has negligible variation (Mehra and Prescott, 1985; Weil, 1989). Habit formation preferences can help to capture the counter-cyclicality of the risk premia (Abel, 1990; Campbell and Cochrane, 1999; Constantinides, 1990) as well as other features of macro-economic outcomes over the business cycle (Jermann, 2010). Bansal and Yaron (2004) combine the preferences of Epstein and Zin (1989) with stochastic volatility of consumption growth and generate time variation in the conditional volatility of the SDF.

Other studies have found different channels for time variation in risk premia, such as differences in risk aversion (Bhamra and Uppal, 2010; Chan and Kogan, 2002; Gomes and Michaelides, 2008); rare disasters (Barro, 2006, 2009; Rietz, 1988; Wachter, 2012); incomplete markets (Constantinides and Duffie, 1996; Gârleanu and Panageas, 2011); participation constraints, (Basak and Cuoco, 1998; Chien, Cole, and Lustig, 2012; Guvenen, 2009); investment shocks (Papanikolaou, 2011) and heterogeneity in the frequency of shocks to fundamentals (Calvet and Fisher, 2007). A brief summary of the aforementioned models and their asset pricing implications are shown in Table 3.4.

The asset pricing implications of the models shown in Table 3.4 provide a general idea of the model-implied variability of Sharpe ratios. Indeed, this variability could be used as a metric to better assess the performance of a model. For example, an asset pricing model with constant Sharpe ratios would fail in explaining the observed variation in empirical Sharpe ratios. On the other hand, a model that predicts highly volatile Sharpe ratios may exceed the true variability observed in the data. Therefore, the variance of Sharpe ratios can be used as a metric to better

assess theoretical asset pricing models. This metric would be in the spirit of the entropy measure recently proposed by Backus, Chernov, and Zin (2012) and studied in Martin (2012).

As a robustness check of the variability generated by OLS methods to calculate Sharpe ratios, I performed a second simulation exercise based on the external habit formation model by Campbell and Cochrane (1999). A brief description of the model and a brief overview of the results are presented below.

### 3.4.4 External Habit Formation Model

In the external habit formation model of Campbell and Cochrane (1999), the consumption dynamics are the same as in the standard Lucas model; that is, consumption growth rates are assumed to be independent and identically distributed. Furthermore, the agent is assumed to have external habit formation preferences. The habit level is assumed to be a slow-moving and heteroscedastic process. The heteroscedasticity of the habit process, the sensitivity function, can be chosen so that the real interest rate in the model is constant or linear in the habit. Further details can be found in Appendix B.5.

I use the same calibrated monthly parameters as those in Campbell and Cochrane (1999) to simulate returns from the model and compute annual expected returns, volatilities and Sharpe ratios using standard OLS techniques. I compare these results with the numerical values implied by the model. Consistent with my previous results, I find that the Sharpe ratios based on standard OLS methods are at least twice more volatile than the model-implied variability. The results are plotted in Figure 3.4. Panel A displays the Sharpe ratio estimates based on OLS methods, while Panel B displays the values of the true Sharpe ratios. Clearly, the Sharpe ratio estimates based on OLS methods are more volatile than the values implied by the habit formation model.

## 3.5 Sharpe Ratios Estimated from Reduced Form Models

The use of data simulated by means of theoretical models helps to better identify the economic reasons that drive the time-series variation. An alternative form of analyzing returns is via reduced form models, which are statistical models that do not impose any economic structure. These models aim to better describe historical data. Moreover, they do not rely on arbitrary predictors and are not subject to the effects of errors in variables or misspecification.

In this section, I introduce the nonlinear latent VAR representation proposed in Brandt and Kang (2004), in which the first and second conditional moments are considered latent variables identified from the history of returns. In this setup, the Sharpe ratio and its dynamics are obtained endogenously as the ratio of the conditional moments of excess returns. The framework is general enough and can be extended to a setup that includes flexible correlation structures and exogenous predictors.

### 3.5.1 Brandt and Kang (2004)

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

$$y_t = \mu_{t-1} + \lambda_{t-1}\varepsilon_t \text{ with } \varepsilon_t \sim \mathcal{N}(0,1) \tag{3.21}$$

where $\mu_{t-1}$and $\lambda_{t-1}$ 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:

$$\begin{bmatrix} \ln \mu_t \\ \ln \lambda_t \end{bmatrix} = d + A \begin{bmatrix} \ln \mu_{t-1} \\ \ln \lambda_{t-1} \end{bmatrix} + \eta_t \text{ with } \eta_t \equiv \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} \sim \mathcal{N}(0, \Sigma), \tag{3.22}$$

where

$$\begin{aligned} d &= \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}, \; A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ and} \\ \Sigma &= \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \text{ with } b_{12} = b_{21} = \rho\sqrt{b_{11}b_{22}}. \end{aligned} \tag{3.23}$$

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

$$\mathbb{E}\begin{bmatrix} \ln \mu_t \\ \ln \lambda_t \end{bmatrix} = (I - A)^{-1} d \tag{3.24}$$

and

$$\text{vec}\left(\text{cov}\begin{bmatrix} \ln \mu_t \\ \ln \lambda_t \end{bmatrix}\right) = (I - (A \otimes A))^{-1} \text{vec}(\Sigma) \tag{3.25}$$

where $\otimes$ represents the Kronecker product.

The key elements of the return dynamics presented Eq. (3.22) are the transition matrix $A$ and the correlation coefficient $\rho$. 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. The correlation coefficient $\rho$ captures the contemporaneous correlation between the innovations to the conditional moments. This parameter is of considerable importance since it captures the risk-return trade-off.[57]

[57]Most asset pricing models predict a positive relationship between the market's risk premium and conditional volatility (Merton, 1973). However, the empirical evidence on the sign of the risk-return relation is inconclusive. Indeed, some studies find a positive relation (Ghysels, Santa-Clara, and Valkanov, 2005; Ludvigson and Ng, 2007; Lundblad, 2007; Pastor, Sinha, and Swaminathan, 2008; Scruggs, 1998), but others find a negative relation (Brandt and Kang, 2004; Campbell, 1987; Glosten, Jagannathan, and Runkle, 1993; Harvey, 2001; Lettau and Ludvigson, 2010). Others have shown through theoretical studies that the intertemporal mean-variance relationship may not be positive or negative (Ang and Liu, 2007; Whitelaw, 2000).

The model in Eq. (3.22) generalizes the permanent and temporary components of Fama and French (1988b) and the standard stochastic volatility model. The equation for the conditional mean is

$$\ln \mu_t = d_1 + a_{11} \ln \mu_{t-1} + a_{12} \ln \lambda_{t-1} + \eta_{1t}, \quad \text{where} \quad \eta_{1t} \sim \mathcal{N}(0, b_{11}). \tag{3.26}$$

If $a_{12} = 0$, the dynamics of the conditional mean is similar to the temporary component as in Lamoureux and Zhou (1996). Now, the equation that describes the conditional volatility is

$$\ln \lambda_t = d_2 + a_{21} \ln \mu_{t-1} + a_{22} \ln \lambda_{t-1} + \eta_{2t}, \quad \text{where} \quad \eta_{2t} \sim \mathcal{N}(0, b_{22}), \tag{3.27}$$

and corresponds to the standard stochastic volatility model; in particular if $a_{21} = 0$, Eq. (3.27) is the standard stochastic volatility model as in Andersen and Sørensen (1996) and Kim, Shephard, and Chib (1998). Finally, we learn from Eq. (3.25) that the unconditional variance is determined by the variance–covariance matrix $\Sigma$ and the matrix $A$. For identification purposes, I assume four different specifications for the transition matrix $A$. First, in model A, I consider the case in which the conditional mean and volatility evolve as in Eqs. (3.26) and (3.27). Models B and C consider $a_{12} = 0$ and $a_{21} = 0$, respectively, allowing for the model of permanent and temporary component in the first case, and the standard stochastic volatility model in the second case. Finally, model D considers the case in which $a_{12} = a_{21} = 0$.

An interesting property is the nonnegativity of expected returns and volatilities. This nonnegativity of the conditional mean guarantees a positive risk premium, as suggested in Merton (1980), and has been used by Bekaert and Harvey (1995) and Jacquier, Johannes, and Polson (2007), among others. The log-normality specification for the volatility is consistent with Andersen, Bollerslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, Diebold, and Labys (2003), which show that the log-volatility process can be well approximated by a normal distribution.

### 3.5.2 Implied Sharpe Ratio

The latent VAR implies a conditional Sharpe ratio of the form

$$SR_t = \frac{\mu_t + \lambda_t^2/2}{\lambda_t}, \tag{3.28}$$

where $\mu_t$ and $\lambda_t$ are the conditional mean and volatility of stock returns.[58] It follows that the Sharpe ratio is stochastic if the innovations that affect both the numerator and denominator in Eq. (3.28) are stochastic and do not cancel each other out. Moreover, the Sharpe ratio is time-varying due to the mean reversion of the two conditional moments. The distribution of the Sharpe ratio corresponds to the sum of two correlated log-normal distributions, which is not standard.

[58] The squared term in the numerator comes from a Jensen's adjustment for log-returns.

### 3.5.3 The Data

I study quarterly returns on the value-weighted index market portfolio from CRSP. Excess returns are calculated by subtracting the quarterly yield on a three-month T-bill from the corresponding stock return. I use this yield instead of the monthly yield due to the idiosyncratic variation documented in Duffee (1996). The predictors are the CRSP dividend-price ratio (*d-p*), calculated as the log-ratio of the CRSP dividends to the price level of the CRSP value-weighted stock index; the relative bill rate ($RREL$), which is the difference between the three-month treasury bill and its four-quarter moving average; the term spread ($TRM$), the difference between the ten-year treasury bond yield and the three-month treasury bill; the default spread ($DEF$), the difference between the BAA corporate bond rate and the AAA corporate bond rate and the consumption-wealth ratio proxy (*cay*).[59] The $RREL$, $TRM$ and $DEF$ are obtained from the Federal Reserve statistical release. Data on the dividend-price ratio is taken from CRSP and the time series of *cay* is taken from Sidney Ludvigson's website.[60] All data are quarterly from the period April 1953 to December 2011.

### 3.5.4 Parameter Estimates

The model in Eqs. (3.21) and (3.22) is nonlinear since the first equation is nonlinear in the state-variables. The parameters are estimated via QML methods and are shown in Table 3.5. The first column corresponds to the estimates of model A, the second column shows the estimates for model B, and the third and fourth columns contain the parameter estimates for models C and D, respectively. Given the frequency of returns, expected returns are persistent since the estimates for $a_{11}$ range from 0.6727 to 0.7204.[61] The conditional volatility is more persistent than the conditional mean, for all model specifications.

The parameter estimates of the models A through D show evidence of a strong and negative risk-return trade-off, measured by the correlation between the innovations to the conditional mean and the volatility of excess returns. The estimates range from -0.1760 to -0.7995, for both the constrained and unconstrained representations, and are statistically significant. This finding is consistent with the negative risk-return relationship found in Brandt and Kang (2004), Campbell and Hentschel (1992b) and Campbell (1987). The negative sign of the correlation coefficient between the conditional mean and the volatility of returns amplifies the variability of the Sharpe ratio, whereas a positive correlation between expected returns and volatility makes Sharpe ratios less variable than its mean or even constant.

[59] These predictors are used in the predictability literature. See Goyal and Welch (2008) and Lettau and Ludvigson (2010) for details.

[60] I thank Sidney Ludvigson for making the time series data for cay available. This variable is calculated in a quarterly basis.

[61] These values correspond to a monthly persistence of roughly 0.87 to 0.89.

The estimates show that there is more variation in the mean than in the conditional volatility, since the conditional variance of the innovations to the conditional mean, $b_{11}$, differs substantially from that of the innovation to the conditional volatility, $b_{22}$. The off-diagonal elements of the transition matrix A are significant. However, the values for $a_{21}$ are similar across models, while the values for $a_{12}$ differ. The differences in signs of $a_{12}$ and $a_{21}$ are consistent with the results of Whitelaw (1994) and Brandt and Kang (2004), which state that the cross-autocorrelations between the conditional mean and volatility offset each other through time.

### 3.5.5 Expected Returns, Volatilities and Sharpe Ratios

Given the parameter estimates in Table 3.5, I estimate expected returns, volatilities and Sharpe ratios via particle filtering. The left column of Figure 3.5 presents the filtered estimates of quarterly expected returns (first row), volatility (second row) and Sharpe ratios (third row). Each plot also shows in vertical bars the NBER recession dates. It is clear that the conditional mean, volatility and Sharpe ratio are time varying. The quarterly mean has a standard deviation of less than 1% and it varies from 1% in the third quarter of 1974 to 3% in the last quarter of 2003. The quarterly volatility has a standard deviation of 2% and ranges from 7.3% to 11.6%. Expected returns revert more quickly to their unconditional mean than do conditional volatilities, and this is consistent with the estimates of $a_{11}$ and $a_{22}$.

Quarterly Sharpe ratios are displayed in the last row of the first column. The Sharpe ratio rises from the peak to the trough of the recession dates in the sample, and is consistent with the empirical results documented by Lustig and Verdelhan (2012); Tang and Whitelaw (2011) and Lettau and Ludvigson (2010). This countercyclical variation of the Sharpe ratio is also consistent with the habit formation models (Campbell and Cochrane, 1999; Constantinides, 1990). Intuitively, at the peak of the business cycle, consumers enjoy consumption levels far above their habits. As a result, a low Sharpe ratio, or low reward per unit of risk, is required for a consumer to invest in the stock index at the peak of the cycle, in contrast to the trough of a cycle, where consumption levels are closer to those of the habits, which makes consumers more relative risk averse. For an investor willing to invest in the trough of the cycle, the rewards per unit of risk or Sharpe ratios should be substantially high.

#### OLS estimates

I calculate expected returns, volatilities and Sharpe ratios based on OLS techniques for comparison purposes. Table 3.6 presents the estimates from OLS regressions of quarterly realized excess returns and excess log-returns from the first quarter of 1953 to the last quarter of 2011. The results are generally consistent with those reported in the predictability literature. There is no substantial difference between the regression estimates obtained by using excess returns and excess log-returns. At a one-quarter horizon, *cay* and $RREL$ show a consistent predictive power for excess returns. Indeed, *cay* alone explains 3% of next quarter's total variability. Adding the

lagged value of excess returns, $cay$, $d-p$, $RREL$ and $TRM$ explains 8% of the quarter's variation in the next quarter's excess return. The $R-$squared of 8% for log-returns is lower than the values reported in previous studies, but the sample, which includes the 2007-2008 financial crisis, may account for this result. The results for the volatility equation are presented in Table 3.7. In this representation, the lagged volatility, $d-p$, $TRM$ and $DEF$ are significant. The positive serial correlation in realized volatility reflects the autoregressive conditional heteroskedasticity of quarterly returns. The lagged value of volatility alone explains 37% of next the quarter's excess return volatility. Lagged volatility values, $cay$, $d-p$, $RREL$, and $TRM$ explain altogether 41%. Finally, the high $R-$squared value of 43% in the full volatility equation reflects that realized volatility is much more predictable than excess returns.

Empirical moments of expected returns, volatilities and Sharpe ratios are displayed in Table 3.8. The first set of estimates is calculated based on OLS regressions of quarterly realized log-returns for the CRSP value-weighted index on lagged explanatory variables. The second set of estimates is based on the reduced form model by Brandt and Kang (2004), in which the conditional mean and volatility of stock returns are treated as latent variables. This representation guarantees positive values for expected returns and volatilities. As in the simulation exercises described in section 4, I find differences worth noting among the estimates. First, expected returns and volatilities calculated via OLS have a quarterly standard deviation of 2%, while the standard deviation of the filtered estimates is 1%. Filtered volatilities are higher, on average, than the ones obtained with OLS methods and more autocorrelated. The autocorrelation of expected returns obtained with OLS methods is 81%, in contrast with the one estimated from the filtered series, which is less than 59%. This is not surprising, since the regressors used for its estimation are highly persistent. The autocorrelation of the filtered estimates is consistent with the estimated value of $a_{11}$.

As for the Sharpe ratio estimates, there are major differences worth noting. First, the average quarterly Sharpe ratio estimated via filtering is 26% while the OLS estimate is 30%. As for the standard deviation estimates, the difference is quite substantial. For the OLS estimates, the standard deviation is 42%, which is similar to the 45% reported by Lettau and Ludvigson (2010), while the standard deviation of the filtered Sharpe ratio ranges from 5%. An explanation of this difference is the use of standard OLS techniques for its estimation. Reduced form representations do not rely on predetermined conditioning variables to estimate conditional moments; the state variables are identified from the history of returns whereas standard OLS methods generate fitted moments from a set of predictive regressions as proxies for the unobservable conditional mean and volatility. The fitted moments depend on the joint distribution of the predictors; therefore, any model misspecification generates fitted moments that do not correspond to the true dynamics of the conditional mean and volatility, and as a result, the Sharpe ratio dynamics. Another important issue is that the ratio of the fitted moments does not adjust for the correlation between expected returns and volatility of stock returns, whereas filtering techniques do.

**Alternative Reduced Forms**

For comparison purposes, I also analyze an unconstrained version of the representation of Brandt and Kang (2004). The excess returns have time-series dynamics of the form

$$y_t = \mu_{t-1} + \lambda_{t-1}\varepsilon_t \text{ with } \varepsilon_t \sim \mathcal{N}(0,1), \tag{3.29}$$

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

$$\begin{bmatrix} \mu_t \\ \ln \lambda_t \end{bmatrix} = d + A \begin{bmatrix} \mu_{t-1} \\ \ln \lambda_{t-1} \end{bmatrix} + \eta_t \text{ with } \eta_t \sim \mathcal{N}(0,\Sigma), \tag{3.30}$$

where $d$, $A$ and $\Sigma$ are defined as in Eq. (3.23). The main difference between the model representation by Brandt and Kang (2004) and Eqs. (3.29) and (3.30) is that expected returns can potentially be negative, as in Lamoureux and Zhou (1996). As in the previous model, I consider four model specifications for the matrix $A$. The covariance matrix, $\Sigma$, has the same structure as Eq. (3.23). The sign of the correlation coefficient between the conditional mean and the volatility of excess returns has the same sign as the correlation between the conditional mean and the log-volatility.[62]

QML estimates of the model with an unconstrained risk premia are shown in Table 3.9. Under all model specifications, the parameter estimates, are similar to the estimates of the first model. An important difference is that the estimates of the off-diagonal elements $a_{12}$ and $a_{21}$ are negative, although $a_{12}$ is not statistically significant. The right column of Figure 3.5 displays the filtered estimates of conditional moments and Sharpe ratios for the model with an unconstrained risk premia. The main difference between the constrained and unconstrained representations is that expected returns can take negative values; indeed, expected return estimates took negative values for six quarters of the whole sample. Qualitatively, both latent VAR models show similar dynamic behavior; in fact, the correlation coefficient between the implied Sharpe ratio estimates is 70%.

**Exogenous Predictors**

The main advantage of the latent VAR approach by Brandt and Kang (2004) is that it allows the study of the dynamics of the conditional mean, volatility and Sharpe ratios without relying on exogenous predictors. At the same time, useful information is potentially discarded, since any correlation structure between predictors and conditional moments is ignored. As a robustness

[62]From Stein's lemma, we have that the conditional covariance between excess returns and the conditional volatility is $cov_{t-1}(\mu_t, \lambda_t) = \mathbb{E}_t[\lambda_t] \cdot cov_{t-1}(\mu_t, \ln \lambda_t)$. Thus, the sign of the correlation coefficient between the conditional mean and the volatility of stock returns is the same as the conditional correlation of the conditional mean and the log-volatility of returns.

check, I estimate an extended version of the model in which each moment is a function of the same exogenous predictors used in the predictive regressions ( $cay$, $d-p$, $RREL$, and $TRM$). The model specification is given by

$$y_t = \mu_{t-1} + \lambda_{t-1}\varepsilon_t \text{ with } \varepsilon_t \sim \mathcal{N}(0,1), \tag{3.31}$$

where

$$\begin{bmatrix} \ln \mu_t \\ \ln \sigma_t \end{bmatrix} = d + A \begin{bmatrix} \ln \mu_{t-1} \\ \ln \sigma_{t-1} \end{bmatrix} + Cx_{t-1} + \eta_t, \text{ with } \eta_t \sim \mathcal{N}(0, \Sigma), \tag{3.32}$$

where $x_t$ denotes the de-meaned vector of predictors observed at date $t$.

Table 3.10 reports the parameter estimates of the extended model D and also replicates for comparison the results of model D. The estimates of $A$ and $\Sigma$ are similar across the two models. When I add the exogenous predictors, all the parameter estimates of the base model decrease in magnitude, which means that the exogenous predictors help explain some of the variation in the moments that was left unexplained. Finally, the correlation between the innovations to the mean and volatility decreases in magnitude from -0.7995 to -0.4523, both significant.

In the mean equation of the extended model, the coefficients of $cay$, $d-p$, $TRM$ ($c_{11}, c_{12}$ and $c_{14}$) are positive and the coefficients of $RREL$ and $DEF$ ($c_{13}$ and $c_{15}$) are negative. In the volatility equation, all coefficients are negative, except for one, $DEF$. The signs of the coefficients are all consistent with the results of the predictive regressions. However, it is important to note that these results are not directly comparable to standard predictive regressions, since these coefficients correspond to regressions with the conditional moments as dependent variables.

**Comparison**

Empirical moments of the different Sharpe ratio estimates are displayed in Table 3.11. The first, second and third sets of Sharpe ratio estimates are based on the latent VAR approach from the model representation by Brandt and Kang (2004). The first representation is based on Eqs. (3.21) and (3.22), while the second representation guarantees a positive volatility only and is based on Eqs. (3.29) and (3.30). The third representation is an extended version of the first model in which the conditional moments are positive functions of exogenous predictors and is represented in Eqs. (3.31) and (3.32). Finally, the last set of Sharpe ratio estimates is based on the conditional moments calculated from OLS regressions of log-returns on lagged explanatory variables.

The results from Table 3.11 show that the average quarterly Sharpe ratios under the first two models are 25% and 26%, respectively. The third model implies a quarterly Sharpe ratio of 31%, while the estimates obtained from OLS methods have a quarterly Sharpe ratio of 30%. The difference is caused by the set of exogenous predictors included within the estimation procedure. The first set of results represents the Sharpe ratio estimates based on the set of observed returns,

while the third and fourth correspond to Sharpe ratio estimates using the history of returns and the set of exogenous predictors. The parameter estimates used in the filtering calculations depend on the data used in the estimation process. In the first two models, the parameter and filtered estimates depend on the time series of excess returns, while the last two models depend on the same series of returns as well as on the set of exogenous predictors.

As for the Sharpe ratio volatility implied by the models, there are some differences worth noting. The first two models imply a volatility of 5% and 10%, respectively. The difference is due to the model representation. The first model considers a positive risk premia and the second does not. Since the second model allows for negative Sharpe ratios, there is more variability. As for the third representation, the variability is 25%, which is mainly driven by the inclusion of a set of exogenous predictors that affect the conditional mean and volatility of excess returns. None of these representations deliver a Sharpe ratio volatility of 42% as OLS methods do. The main driver of this difference is the use of conditioning information within the estimation process. In the first two cases, the model representations as well as the history of returns determine the variability of the Sharpe ratio. In contrast, the set of exogenous predictors that are included in the estimation process of the third model and fourth model determines a higher variability of the Sharpe ratio estimates.

## 3.6 Implications for Portfolio Choice

In this section, I discuss a standard model from the portfolio-choice literature and its relation to the market Sharpe ratio.

### 3.6.1 Portfolio Optimization: One Risky Asset

I consider an investor with mean-variance preferences that optimizes the tradeoff between the mean and the variance of portfolio returns. Two assets are available to an investor at time $t$. One is risk free, with return $R_{f,t+1}$ from time $t$ to time $t+1$, and the other is risky. The risky asset has simple return $R_{t+1}$ from time $t$ to time $t+1$ with conditional mean $\mathbb{E}_t[R_{t+1}]$ and conditional variance $\sigma_t^2$. The investor allocates a share $\alpha_t$ of her portfolio into the risky asset. Then the portfolio return is

$$R_{p,t+1} = R_{f,t+1} + \alpha_t \left(R_{t+1} - R_{f,t+1}\right).$$

The mean portfolio return is $\mathbb{E}_t[R_{p,t+1}] = R_{f,t+1} + \alpha_t \left(\mathbb{E}_t[R_{t+1}] - R_{f,t+1}\right)$, while the variance of the portfolio is $\sigma_{pt}^2 = \alpha_t^2 \sigma_t^2$. If the investor has mean-variance preferences, then she trades off between the mean and variance in a linear fashion. In other words, she maximizes a linear combination of mean and variance with a positive weight on mean and a negative weight on variance,

$$\max_{\alpha_t} \left(\mathbb{E}_t[R_{p,t+1}] - \frac{\gamma}{2}\sigma_{pt}^2\right).$$

The solution to this optimization problem is

$$\alpha_t = \frac{\mathbb{E}_t\left[R_{t+1}\right] - R_{f,t+1}}{\gamma\sigma_t^2}. \tag{3.33}$$

The optimal weight for the stock index coincides with the myopic demand and can be interpreted as the product of the relative risk tolerance[63] and the market Sharpe ratio normalized by the volatility of the market returns; that is,

$$\alpha_t = \frac{SR_t}{\gamma\sigma_t}. \tag{3.34}$$

We learn from Eq. (3.34) that for investors with mean-variance preferences, the optimal allocation in the market portfolio is determined by three elements: the Sharpe ratio of the market portfolio, the conditional volatility of the market portfolio and the risk aversion parameter. Moreover, the variability of portfolio weights is determined by the variability of Sharpe ratios and the standard deviation of the market portfolio.

Campbell and Viceira (2002) derive a similar expression by assuming an investor with power utility and that the return on an investor's portfolio is lognormal, with the slight difference that the optimal weight in Eq. (3.33) is adjusted by half the variance of the risky asset; that is,

$$\alpha_t = \frac{\mathbb{E}_t\left[r_{t+1}\right] - r_{f,t+1} + \sigma_t^2/2}{\gamma\sigma_t^2}. \tag{3.35}$$

Now I implement the model following the standard plug-in approach; that is, I solve the optimization problem assuming that the mean and variance of returns are known. Once the problem is solved, I replace the moments with their estimates obtained via regression or filtering techniques. For simplicity, I assume that the investor ignores estimation risk while making an investment decision.

Figure 3.6 shows the optimal allocations in Eq. (3.35) using OLS and filtering methods to estimate conditional moments assuming a risk aversion parameter $\gamma = 5$. Clearly, the portfolio weights constructed via OLS methods are more volatile than the ones obtained with the filtered moments. Indeed, the average portfolio weight under the OLS model specification is 1.27 with a standard deviation of 2.13, in contrast to the portfolio weight computed with filtering methods, which is on average 56% with a standard deviation of 12%. Finally, the correlation between the two weights is 15%. These results have practical implications for portfolio allocation, especially for an investor who faces proportional costs by trading the optimal portfolio of an investor with mean-variance preferences.[64] As the optimal weight is proportional to the market Sharpe ratio, the percentage of wealth traded in each period will depend upon the volatility of the market

[63] This term is the inverse of the relative risk aversion.

[64] This fact was noted by De Miguel, Garlappi, and Uppal (2009) for performance evaluation.

Sharpe ratio. It is clear that upward-biased estimates of the Sharpe ratio volatility would imply excessive portfolio rebalancing, and therefore more transaction costs.

## 3.7 Concluding Remarks

In this chapter I examine whether estimates of the variability of the Sharpe ratio may be biased due to limitations of the empirical methodology used in its estimation. I provide evidence that measurement error in estimated Sharpe ratios helps to explain the Sharpe ratio volatility puzzle. I further show that this measurement error is caused by the use of standard OLS methods to estimate the ratio. The empirical question I address is important because many studies have used the results implied by OLS methods to calibrate the volatility of the market Sharpe ratio.

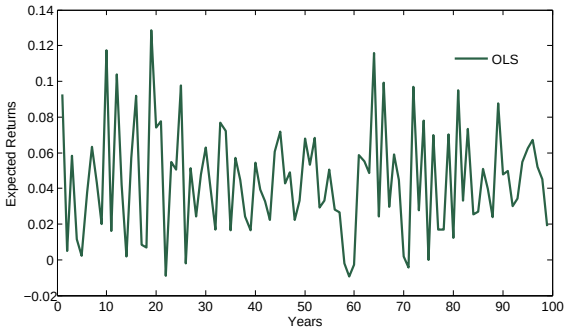
Based on simulated data from standard asset pricing models, I document that OLS methods produce Sharpe ratio volatility that is larger than the ratio's true variability. Using the OLS approach may also provide conditional moment estimates that do not necessarily correspond to their true values.

Once I have documented the upward bias in the Sharpe ratio's variability generated by OLS methods, I consider if using improved empirical methodologies may better reflect the ratio's true variability. To accomplish this goal, I propose filtering methods as a way to better assess this variation. These techniques explicitly allow for the estimation of time-varying moments by aggregating the entire history of realized returns in a parsimonious way. Moreover, filtering is flexible enough to be used with large information sets without relying on exogenous predictors, while being robust to structural breaks. I also show that filtering techniques better reflect the true variation of Sharpe ratios even when parameter values need to be estimated.

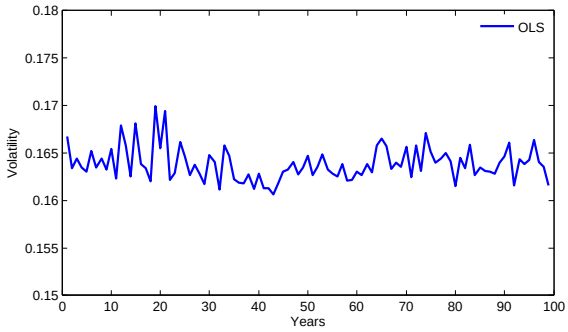
Motivated by the simulation results, I use real data on excess stock returns to compare the Sharpe ratio volatility estimates produced by OLS and filtering methods. I find that filtering methods deliver Sharpe ratio variability estimates that are much smaller than the Sharpe ratio variability estimates implied from OLS methods. The difference in results from the two methodologies arises due to the use of conditioning information within the filtering estimation process.

My findings have significant implications for asset pricing. For example, in a portfolio allocation setting, the optimal portfolio weight is proportional to the market Sharpe ratio. Thus, upward biased estimates of the Sharpe ratio volatility would imply excessive portfolio rebalancing, and therefore more transaction costs.

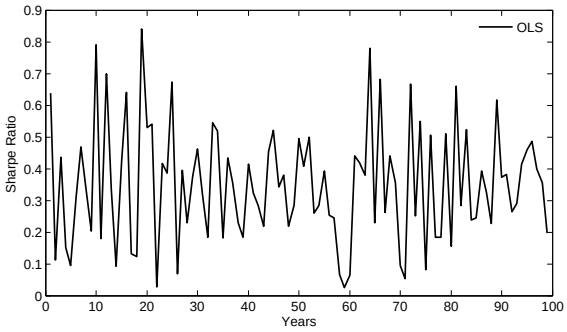
## 3.8 Figures and Tables



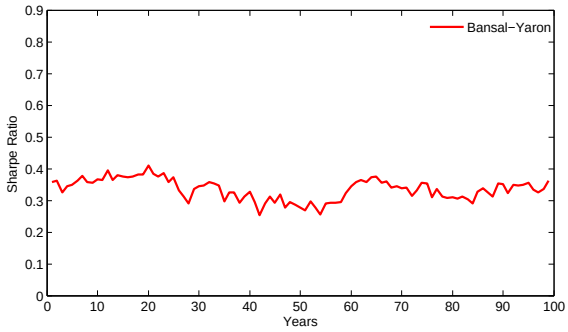

**(a)** Expected Returns (OLS)




**(b)** Conditional Volatility (OLS)



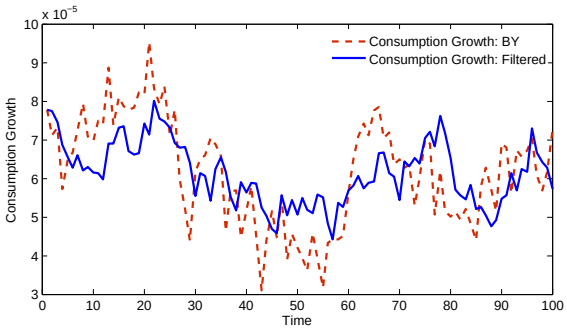

**(c)** Conditional Sharpe Ratios (OLS)



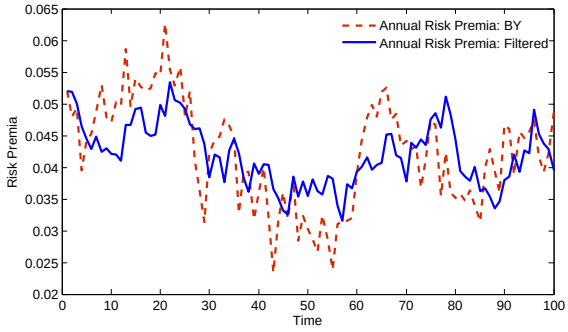

**(d)** Conditional Sharpe Ratios (BY)

**Figure 3.1.** Comparison OLS Estimates versus Simulated Values: Long-Run Risk Model

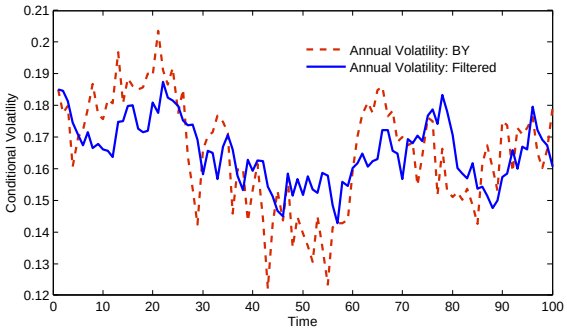
This figure shows the results of a simulated path of annual returns using the calibration by Bansal and Yaron (2004). Each simulation has 100 annual observations of returns. Fitted values for the conditional mean and variance were constructed using predictor variables. Panel A shows a random path of annual returns with the fitted OLS values. Panel B shows the realized variance constructed with realized returns along with its OLS fitted values in dotted lines. Panel C contains the conditional Sharpe ratio estimates based on the OLS fitted values of the conditional mean and conditional volatility; Panel D contains the Sharpe ratios implied by the model.



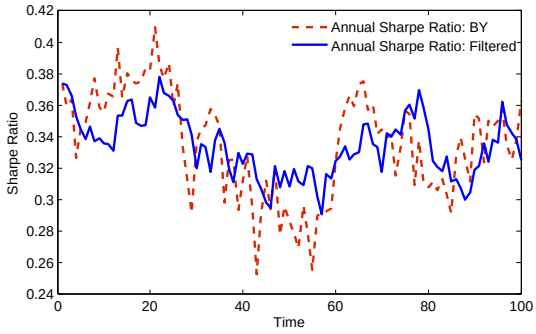

**(a)** Consumption Growth




**(b)** Risk Premia



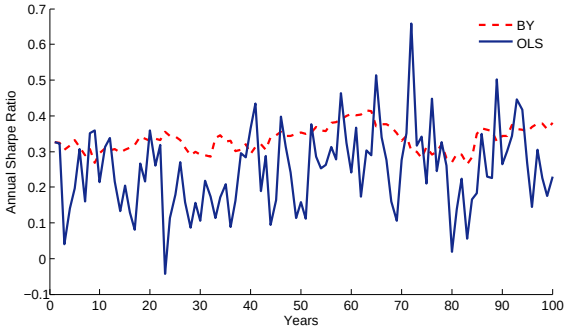

**(c)** Conditional Volatility



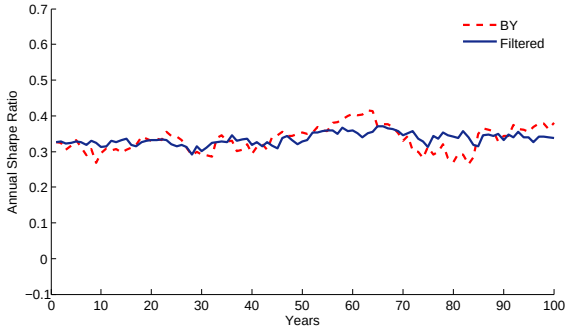

**(d)** Conditional Sharpe Ratio

**Figure 3.2.** Comparison Simulations versus Filtered values: Long-Run Risks Model

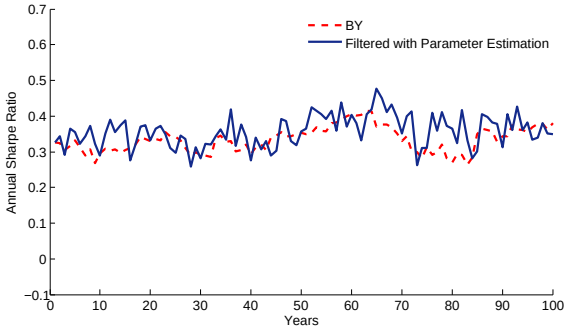
This figure shows the results of a simulated path of the volatility of consumption growth using the calibration by Bansal and Yaron (2004). Each simulation has 100 annual return observations. Panel A shows a random path of monthly returns of the volatility of consumption growth. The dotted line represents the filtered values of $\sigma_t^2$. Panel B shows the simulated risk premia along with its filtered values in dotted lines. Panel C contains the simulated standard deviation of the risk premia as well as its filtered values. Panel D contains the simulated conditional Sharpe ratio along with its filtered values. The dashed lines are assumed to be unobservable to the econometrician, while the continuous lines are the filtered values.




**(a)** OLS Fitted Values



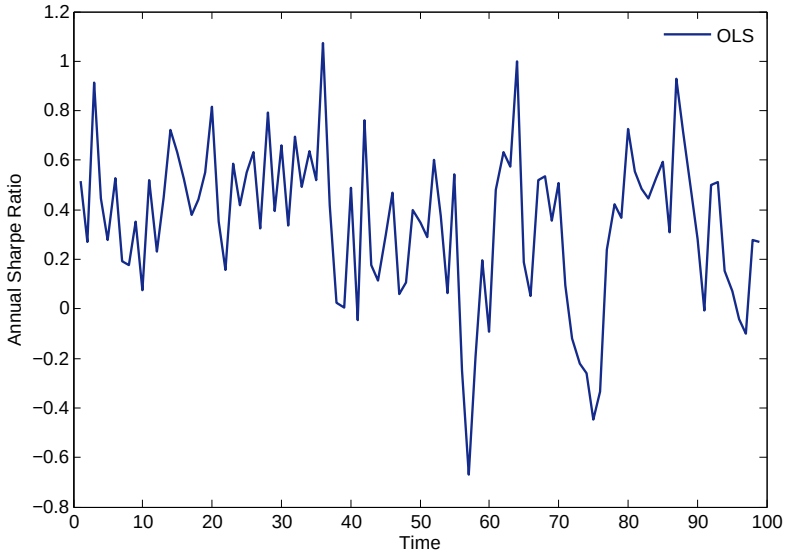

**(b)** Filtered Values (known parameter values)



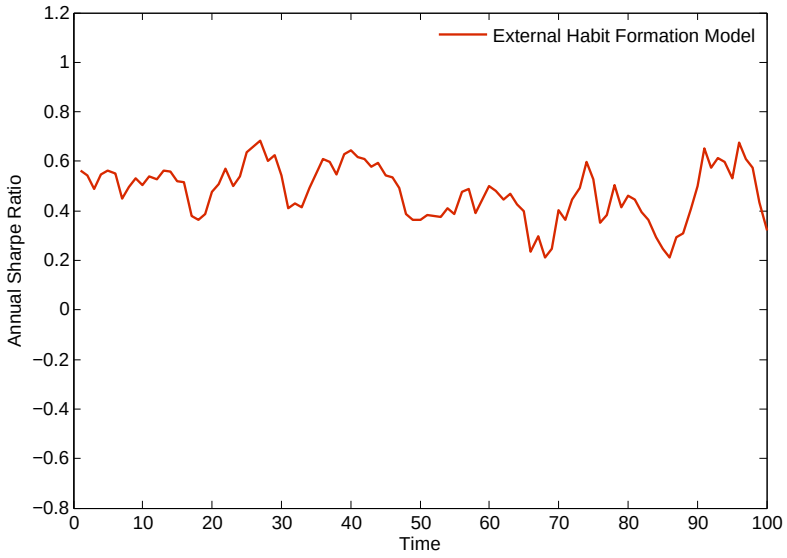

**(c)** Filtered Values (unknown parameter values)

**Figure 3.3.** Comparison OLS versus Filtered Estimates

This figure shows the results of a simulated path of the volatility of consumption growth using the calibration by Bansal and Yaron (2004). Each simulation has 100 annual return observations. Panel A contains the conditional Sharpe ratio estimates based on the OLS fitted values. Panel B contains the filtered Sharpe ratio estimates implied by the long-run risks model; the dotted lines represent the annual Sharpe ratio implied by the model which are assumed to be unobservable to the econometrician; Panel C contains the filtered Sharpe ratio estimates implied by the long-run risks model based on parameter estimates obtained via QML. The dotted lines represent the annual Sharpe ratio implied by the model, which are assumed to be unobservable to the econometrician. The simulations were performed with the calibrated parameter values from Bansal and Yaron (2004).




**(a)** OLS




**(b)** External Habit Formation

**Figure 3.4.** Comparison OLS Estimates versus Simulated Values: Habit Formation Model

This figure shows the results of a simulated path of the volatility of consumption growth using the calibration by Campbell and Cochrane (1999). Each simulation has 100 annual return observations. Panel A contains the conditional Sharpe ratio estimates based on the OLS fitted values; Panel B contains the filtered Sharpe ratio estimates implied by the external habit formation model. The simulations were performed with the calibrated parameter values from Campbell and Cochrane (1999).

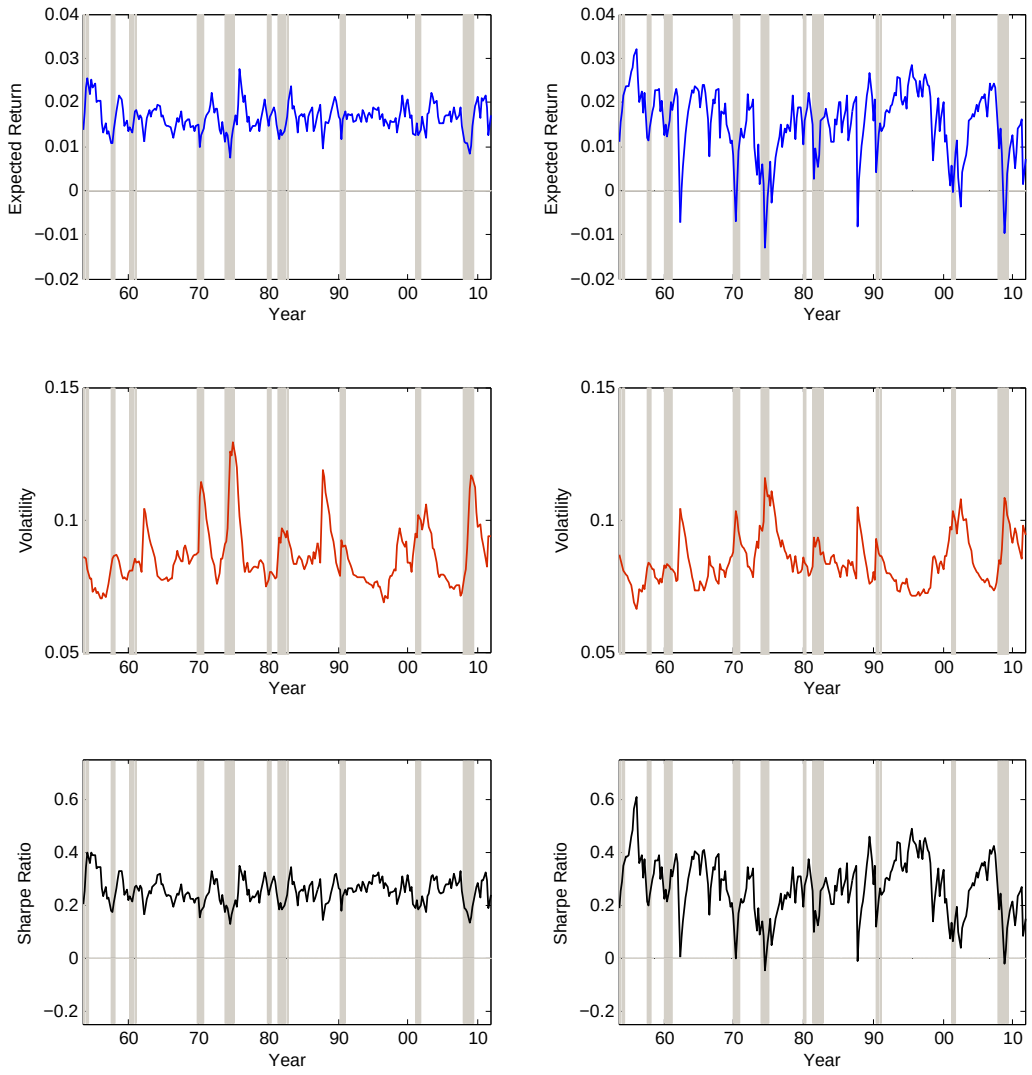



**Figure 3.5.** Expected Returns, Volatility and Sharpe Ratio Estimates. This figure shows the conditional mean, volatility and Sharpe ratio estimates. The figures show the quarterly estimates of the conditional mean, $\mu_t$, conditional volatility, $\sigma_t$ and Sharpe ratio, $SR_t$, obtained via filtering techniques. The left column corresponds to the model with a positive risk premia and the right column contains the filtered estimates of the model with an unconstrained risk premia. The vertical bars represent the NBER recession dates.

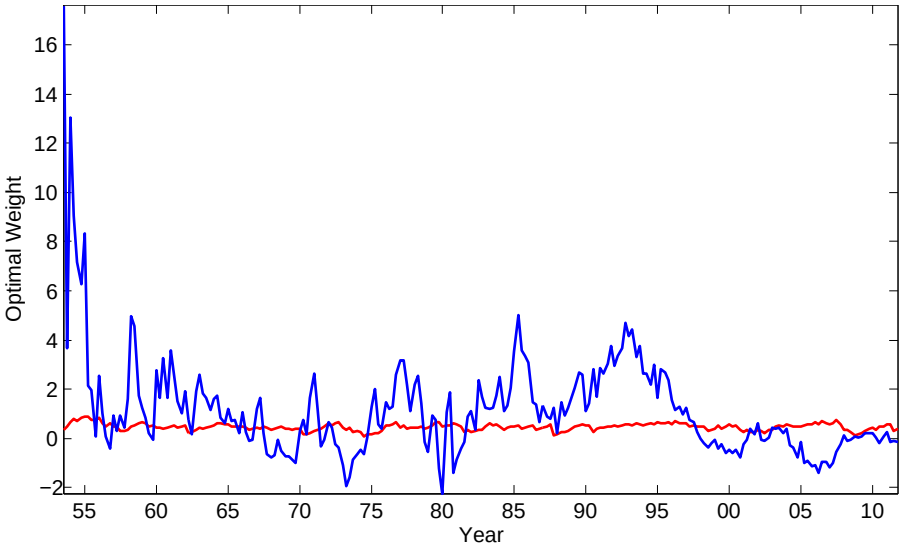



**Figure 3.6.** Portfolio Weights

This figure shows the portfolio weights estimates based on the conditional mean, volatility and Sharpe ratio. The figure shows the time series of optimal weights, $w_t = (\mu_t + \sigma_t^2/2)/(\gamma\sigma_t^2)$, where $\gamma$ represents the risk aversion parameter, and $\mu_t$ and $\sigma_t$ are the quarterly estimates of the conditional mean and conditional volatility respectively. The figure shows the optimal weights based on OLS techniques (blue) and the model based on nonlinear latent variables, assuming a positive risk premium (red) and $\gamma = 5$. The vertical bars represent the NBER recession dates.

**Table 3.1.** Long-Run Risks Parameters

| **Endowment Process Parameters** | **Symbol** | **BY Calibration** |
|---|---|---|
| Mean Consumption Growth | $\mu_c$ | 0.0015 |
| LRR Persistence | $\rho$ | 0.979 |
| LRR Volatility Multiple | $\varphi_e$ | 0.044 |
| Mean Dividend Growth | $\mu_d$ | 0.0015 |
| Dividend Leverage | $\phi$ | 3 |
| Dividend Volatility Multiple | $\varphi$ | 4.5 |
| Dividend Consumption Exposure | $\pi$ | 0 |
| Baseline Volatility | $\overline{\sigma}$ | 0.0078 |
| Volatility of Volatility | $\sigma_w$ | 0.0000023 |
| Persistence of Volatility | $\nu$ | 0.987 |

| **Preference Parameters** | **Symbol** | **BY Calibration** |
|---|---|---|
| Risk Aversion | $\gamma$ | 10 |
| EIS | $\psi$ | 1.5 |
| Time Discount Factor | $\delta$ | 0.998 |

**Endowment Process:**

$$\begin{aligned}
\Delta c_{t+1} &= \mu_c + x_t + \sigma_t \eta_{t+1} \\
x_{t+1} &= \rho x_t + \varphi_e \sigma_t e_{t+1} \\
\sigma^2_{t+1} &= \overline{\sigma}^2 + v\left(\sigma^2_t - \overline{\sigma}^2\right) + \sigma_w w_{t+1} \\
\Delta d_{t+1} &= \mu_d + \phi x_t + \varphi \sigma_t u_{t+1} + \pi \sigma_t \eta_{t+1} \\
w_{t+1}, e_{t+1}, u_{t+1}, \eta_{t+1} &\sim i.i.d.\ \mathcal{N}(0,1).
\end{aligned}$$

**Parameter values.** This table displays the model parameters for Bansal and Yaron (2004) (BY). The endowment process is described above. All parameters are given in monthly terms. The standard deviation of the long-run innovations is equal to the volatility of consumption growth times the long-run volatility multiple, and the standard deviation of dividend growth innovations is equal to the volatility of consumption growth times the volatility multiple for dividend growth. Dividend consumption exposure is the magnitude of the impact of the one-period consumption shock on dividend growth. Dividend leverage is the exposure of dividend growth to long-run risks.

**Table 3.2.** Long-Run Risks Moment Comparison: OLS

| Moment | OLS Regressions | Model |
|---|---|---|
| Expected Returns | 0.0417 | 0.0417 |
| Standard Deviation | 0.0301 | 0.0087 |
| Correlation | 0.0052 | |
| Volatility | 0.1653 | 0.1641 |
| Standard Deviation | 0.0092 | 0.0167 |
| Correlation | 0.0434 | |
| Conditional Sharpe Ratio | 0.2645 | 0.3333 |
| Standard Deviation | 0.1582 | 0.0353 |
| Correlation | 0.0039 | |

**Simulation results.** This table displays moments calculated for the Bansal and Yaron (2004) model from annual data-sets. Columns 1 and 2 display the results using years as time interval. The moment displayed is the median from 100,000 finite sample simulations of length 100 years. The returns on equity and the risk-free rate are aggregated to a yearly level by adding the log-returns within the year.

**Table 3.3.** Long-Run Risks Moment Comparison: Filtering

| Moment | Filtering | Model |
|---|---|---|
| Expected Returns | 0.0418 | 0.0417 |
| Standard Deviation | 0.0080 | 0.0087 |
| Correlation | 0.5721 | |
| Volatility | 0.1645 | 0.1650 |
| Standard Deviation | 0.0168 | 0.0167 |
| Correlation | 0.5651 | |
| Conditional Sharpe Ratio | 0.3341 | 0.3333 |
| Standard Deviation | 0.0322 | 0.0353 |
| Correlation | 0.5694 | |

**Simulation results.** This table displays moments calculated for the Bansal and Yaron (2004) model. Columns 2 to 5 display the results using years as time interval. The moment displayed is the median from 1500 finite sample simulations of length 100 years. The returns on equity and the risk-free rate are aggregated to a yearly level by adding the log-returns within the year.

**Table 3.4.** Asset Pricing Models

| | Preferences | | | Time-Varying | | |
|---|---|---|---|---|---|---|
| | **CRRA** | **Recursive** | **Habit** | **Equity Premium** | **Volatility** | **Sharpe Ratios** |
| Lucas (1978) | ✓ | | | | | |
| Breeden (1979) | ✓ | | | | | |
| Mehra and Prescott (1985) | ✓ | | | ✓ | ✓ | ✓ |
| Rietz (1988) | ✓ | | | ✓ | ✓ | ✓ |
| Weil (1989) | | ✓ | | ✓ | ✓ | ✓ |
| Constantinides (1990) | | | ✓ | | | |
| Abel (1990) | | | ✓ | ✓ | ✓ | |
| Campbell and Cochrane (1999) | | | ✓ | ✓ | ✓ | ✓ |
| Chan and Kogan (2002) | | | ✓ | ✓ | ✓ | ✓ |
| Menzly, Santos, and Veronesi (2004) | | | ✓ | ✓ | ✓ | ✓ |
| Bansal and Yaron (2004) | | ✓ | | ✓ | ✓ | ✓ |
| Barro (2006) | ✓ | | | | | |
| Calvet and Fisher (2007) | | ✓ | | ✓ | ✓ | ✓ |
| Barro (2009) | | ✓ | | | | |
| Jermann (2010) * | | | | ✓ | ✓ | ✓ |
| Papanikolaou (2011) | | ✓ | | ✓ | ✓ | ✓ |
| Wachter (2012) | | ✓ | | ✓ | ✓ | ✓ |
| Chien, Cole, and Lustig (2012) | ✓ | | | ✓ | ✓ | ✓ |

**Asset pricing models.** This table compares features of asset pricing models which have been used to price the aggregate stock market: Lucas (1978), Breeden (1979), Mehra and Prescott (1985), Rietz (1988), Weil (1989), Constantinides (1990), Abel (1990), Campbell and Cochrane (1999), Chan and Kogan (2002), Menzly, Santos, and Veronesi (2004), Bansal and Yaron (2004), Barro (2006), Calvet and Fisher (2007), Barro (2009), Jermann (2010), Papanikolaou (2011), Wachter (2012),Chien, Cole, and Lustig (2012). The comparison table is divided into two panels. The first panel focuses on the features of the model (preferences, endowment and technology), while the second focuses on the pricing implications of the various models.

**Table 3.5.** Quasi-Maximum Likelihood Parameter Estimates

**Positive Risk Premia**

| Parameters | Model A | | Model B | | Model C | | Model D | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | S.E. | Estimate | S.E. | Estimate | S.E. | Estimate | S.E. |
| $a_{11}$ | 0.6727 | 0.0066 | 0.7029 | 0.0041 | 0.7204 | 0.0834 | 0.7079 | 0.1211 |
| $a_{21}$ | -0.0894 | 0.0279 | -0.1521 | 0.0184 | - | - | - | - |
| $a_{12}$ | 0.3215 | 0.0011 | - | - | -0.4948 | 0.0938 | - | - |
| $a_{22}$ | 0.8310 | 0.0025 | 0.7400 | 0.0114 | 0.9182 | 0.1798 | 0.8730 | 0.1142 |
| $b_{11}$ | 0.2897 | 0.0063 | 0.1350 | 0.0194 | 0.0944 | 0.5390 | 0.1924 | 0.1674 |
| $b_{22}$ | 0.0020 | 0.0070 | 0.0001 | 0.0054 | 0.0055 | 0.1111 | 0.0072 | 0.8430 |
| $\rho$ | -0.3073 | 0.0009 | -0.1760 | 0.0004 | -0.7989 | 0.2773 | -0.7995 | 0.0029 |
| $\overline{\mu}$ | 0.0131 | 0.0000 | 0.0131 | 0.0159 | 0.0131 | 0.0675 | 0.0131 | 0.1065 |
| $\overline{\sigma}$ | 0.0857 | 0.0000 | 0.0857 | 0.0005 | 0.0857 | 0.0166 | 0.0857 | 0.5172 |
| $\mathcal{L}$ | 245.37 | | 245.32 | | 244.93 | | 244.69 | |

**Estimation results.** This table presents the QML estimates of the models of the form

$$y_t = \mu(S_{t-1}) + \lambda(S_{t-1})\varepsilon_t,$$

and

$$S_t = AS_{t-1} + \underline{\eta_t} \text{ with } \underline{\eta_t} \sim \mathcal{N}(0, \Sigma),$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \Sigma = \begin{bmatrix} b_{11} & \rho\sqrt{b_{11}b_{22}} \\ \rho\sqrt{b_{11}b_{22}} & b_{22} \end{bmatrix}.$$

$\mu(S_t) = \overline{\mu}\exp(S_{1t})$ and $\sigma(S_t) = \overline{\sigma}\exp(S_{2t})$. The estimates are for quarterly returns on the value-weighted CRSP index in excess of the three-month Treasury bill from the second quarter of 1953 to the fourth quarter of 2011. Standard errors are reported in the column next to the parameter estimate.

**Table 3.6.** Regressions on Quarterly Data

| No. | Constant | Lag | Cay | d-p | RREL | TRM | DEF | $R^2$ |
|---|---|---|---|---|---|---|---|---|
| **Panel A: Excess Returns: 1953:2 - 2011:4** | | | | | | | | |
| | | | **1953:2 - 2011:4** | | | | | |
| 1 | 0.01 | 0.07 | | | | | | 0.01 |
| | (2.76) | (1.18) | | | | | | |
| 2 | 0.02 | | 0.79 | | | | | 0.03 |
| | (2.92) | | (2.57) | | | | | |
| 3 | 0.01 | 0.08 | 0.80 | | | | | 0.03 |
| | (2.81) | (1.28) | (2.71) | | | | | |
| 4 | 0.14 | | 0.76 | 0.02 | | | | 0.04 |
| | (1.84) | | (2.30) | (1.61) | | | | |
| 5 | 0.13 | 0.08 | 0.67 | 0.02 | -1.46 | | | 0.07 |
| | (1.86) | (1.31) | (2.16) | (1.64) | (-2.64) | | | |
| 6 | 0.16 | 0.06 | 0.61 | 0.03 | -1.29 | 0.84 | | 0.08 |
| | (2.12) | (0.99) | (1.99) | (1.96) | (-2.20) | (1.59) | | |
| 7 | 0.16 | 0.06 | 0.61 | 0.03 | -1.30 | 0.84 | -0.07 | 0.08 |
| | (1.87) | (0.99) | (1.99) | (1.83) | (-2.30) | (1.59) | (-0.05) | |
| **Panel B:Log Excess Returns:** | | | | | | | | |
| | | | **1953:2 - 2011:4** | | | | | |
| 1 | 0.01 | 0.08 | | | | | | 0.01 |
| | (2.14) | (1.31) | | | | | | |
| 2 | 0.01 | | 0.82 | | | | | 0.03 |
| | (2.34) | | (2.65) | | | | | |
| 3 | 0.01 | 0.09 | 0.83 | | | | | 0.03 |
| | (2.18) | (1.41) | (2.81) | | | | | |
| 4 | 0.15 | | 0.78 | 0.03 | | | | 0.04 |
| | (1.96) | | (2.35) | (1.77) | | | | |
| 5 | 0.15 | 0.09 | 0.70 | 0.03 | -1.37 | | | 0.07 |
| | (2.00) | (1.43) | (2.23) | (1.81) | (-2.38) | | | |
| 6 | 0.17 | 0.07 | 0.65 | 0.03 | -1.21 | 0.79 | | 0.08 |
| | (2.25) | (1.14) | (2.07) | (2.11) | (-1.99) | (1.49) | | |
| 7 | 0.18 | 0.07 | 0.63 | 0.03 | -1.26 | 0.83 | -0.38 | 0.08 |
| | (2.07) | (1.14) | (2.01) | (2.04) | (-2.17) | (1.56) | (-0.28) | |

**OLS estimation results.** This table reports estimates from OLS regressions of quarterly realized returns and log-returns for the CRSP VW index on lagged explanatory variables for the second quarter of 1953 to the fourth quarter of 2011. The conditioning variables are lagged realized volatility ($Lag$); the consumption, wealth, income ratio ($cay$); log dividend-price ratio ($d-p$); the relative bill rate ($RREL$); the term spread, the difference between the ten-year Treasury bond yield and the three-month Treasury bond yield ($TRM$); the Baa-Aaa default spread ($DEF$). The t-stats were constructed with heteroscedasticity-consistent standard errors.

**Table 3.7.** Regressions on Quarterly Data

| No | Constant | Lag | Cay | d-p | RREL | TRM | DEF | $R^2$ |
|---|---|---|---|---|---|---|---|---|
| **Panel C: Realized Volatility of Excess Returns: 1953:2 - 2011:4** | | | | | | | | |
| 1 | 0.03 | 0.61 | | | | | | 0.37 |
| | (5.16) | (7.39) | | | | | | |
| 2 | 0.07 | | -0.32 | | | | | 0.02 |
| | (16.76) | | (-2.02) | | | | | |
| 3 | 0.03 | 0.60 | -0.23 | | | | | 0.38 |
| | (5.35) | (7.54) | (-2.76) | | | | | |
| 4 | -0.08 | | -0.28 | -0.03 | | | | 0.12 |
| | (-2.16) | | (-1.69) | (-3.90) | | | | |
| 5 | -0.05 | 0.54 | -0.23 | -0.02 | -0.27 | | | 0.41 |
| | (-2.74) | (6.84) | (-2.58) | (-3.97) | (-1.00) | | | |
| 6 | -0.06 | 0.54 | -0.22 | -0.02 | -0.28 | -0.07 | | 0.41 |
| | (-3.02) | (6.84) | (-2.48) | (-4.39) | (-1.06) | (-0.51) | | |
| 7 | -0.10 | 0.46 | -0.16 | -0.02 | -0.18 | -0.19 | 1.34 | 0.43 |
| | (-3.64) | (5.33) | (-1.71) | (-4.57) | (-0.75) | (-1.28) | (2.83) | |
| **Panel D: Realized Volatility of Log Excess Returns: 1953:2 - 2011:4** | | | | | | | | |
| 1 | 0.03 | 0.61 | | | | | | 0.36 |
| | (5.15) | (7.38) | | | | | | |
| 2 | 0.07 | | -0.32 | | | | | 0.02 |
| | (16.7) | | (-2.01) | | | | | |
| 3 | 0.03 | 0.60 | -0.23 | | | | | 0.37 |
| | (5.34) | (7.53) | (-2.75) | | | | | |
| 4 | -0.08 | | -0.28 | -0.03 | | | | 0.12 |
| | (-2.18) | | (-1.68) | (-3.92) | | | | |
| 5 | -0.05 | 0.54 | -0.23 | -0.02 | -0.27 | | | 0.41 |
| | (-2.76) | (6.83) | (-2.57) | (-3.99) | (-0.99) | | | |
| 6 | -0.06 | 0.54 | -0.22 | -0.02 | -0.28 | -0.07 | | 0.41 |
| | (-3.04) | (6.83) | (-2.46) | (-4.41) | (-1.05) | (-0.48) | | |
| 7 | -0.10 | 0.46 | -0.16 | -0.02 | -0.17 | -0.19 | 1.34 | 0.42 |
| | (-3.65) | (5.32) | (-1.71) | (-4.58) | (-0.73) | (-1.24) | (2.81) | |

**OLS estimation results.** This table reports estimates from OLS regressions of quarterly realized volatility of returns and log-returns for the CRSP VW index on lagged explanatory variables for the second quarter of 1953 to the fourth quarter of 2011. The conditioning variables are lagged realized volatility ($Lag$); the consumption, wealth, income ratio ($cay$); log dividend-price ratio ($d-p$); the relative bill rate ($RREL$); the term spread, the difference between the ten-year Treasury bond yield and the three-month Treasury bond yield (TRM); the Baa-Aaa default spread ($DEF$). The t-stats were constructed with heteroscedasticity-consistent standard errors.

**Table 3.8.** Summary Statistics of Expected Returns, Volatilities and Sharpe Ratio Estimates

| | | Mean | Std. Dev. | Min. | Max. | A.C.(1) |
|---|---|---|---|---|---|---|
| OLS Methods | $\mu_t$ | 0.01 | 0.02 | -0.04 | 0.07 | 0.80 |
| | $\sigma_t$ | 0.07 | 0.02 | 0.02 | 0.22 | 0.79 |
| | $SR_t$ | 0.30 | 0.42 | -0.61 | 1.84 | 0.81 |
| Brandt and Kang (2004) | $\mu_t$ | 0.02 | 0.01 | 0.01 | 0.03 | 0.59 |
| | $\sigma_t$ | 0.09 | 0.01 | 0.07 | 0.12 | 0.85 |
| | $SR_t$ | 0.25 | 0.05 | 0.14 | 0.41 | 0.61 |

**Moment comparison.** This table reports descriptive statistics of the estimates of expected returns, volatilities and Sharpe ratios. The first set of conditional moments are estimated from OLS regressions of quarterly realized log-returns for the CRSP VW index on lagged explanatory variables for the first quarter of 1953 to the last quarter of 2011. The second set of estimates are based on the reduced form model by Brandt and Kang (2004) in which the conditional mean and volatility of stock returns are treated as latent variables.

**Table 3.9.** Quasi-Maximum Likelihood Parameter Estimates

**Positive Risk Premia**

| **Parameters** | **Model A** | | **Model B** | | **Model C** | | **Model D** | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | S.E. | Estimate | S.E. | Estimate | S.E. | Estimate | S.E. |
| $a_{11}$ | 0.5276 | 0.0498 | 0.5532 | 0.0002 | 0.5090 | 0.0440 | 0.5282 | 0.0026 |
| $a_{21}$ | -0.4967 | 0.0206 | -0.4154 | 0.0001 | - | - | - | - |
| $a_{12}$ | -0.0165 | 0.1388 | - | - | -0.0247 | 0.3278 | - | - |
| $a_{22}$ | 0.8426 | 0.1521 | 0.8551 | 0.0000 | 0.8859 | 0.9465 | 0.8221 | 0.0068 |
| $b_{11}$ | 0.0002 | 0.0014 | 0.0004 | 0.0012 | 0.0001 | 0.0029 | 0.0004 | 0.0020 |
| $b_{22}$ | 0.0088 | 0.0091 | 0.0048 | 0.0013 | 0.0091 | 0.1097 | 0.0132 | 0.1982 |
| $\rho$ | -0.7994 | 0.0409 | -0.7491 | 0.0002 | -0.7999 | 0.2009 | -0.7678 | 0.0012 |
| $\overline{\mu}$ | 0.0131 | 0.0259 | 0.0131 | 0.0300 | 0.0131 | 0.0013 | 0.0131 | 0.1253 |
| $\overline{\sigma}$ | 0.0857 | 0.0045 | 0.0857 | 0.0036 | 0.0857 | 0.0001 | 0.0857 | 0.0106 |
| $\mathcal{L}$ | 246.12 | | 245.72 | | 246.01 | | 245.52 | |

**Estimation results.** This table presents the QML estimates of the models of the form:

$$y_t = \mu(S_{t-1}) + \lambda_(S_{t-1})\varepsilon_t,$$

and

$$S_t = AS_{t-1} + \underline{\eta_t} \text{ with } \underline{\eta_t} \sim \mathcal{N}(0, \Sigma),$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \Sigma = \begin{bmatrix} b_{11} & \rho\sqrt{b_{11}b_{22}} \\ \rho\sqrt{b_{11}b_{22}} & b_{22} \end{bmatrix},$$
$$\mu(S_t) = \overline{\mu} + S_{1t} \text{ and } \sigma(S_t) = \overline{\sigma}\exp(S_{2t}).$$

The estimates are for quarterly returns on the value-weighted CRSP index in excess of the three-month Treasury bill from the second quarter of 1953 to the fourth quarter of 2011. Standard errors are reported in parentheses.

**Table 3.10.** Quasi-Maximum Likelihood Parameter Estimates: Model with Predictors

| Parameters | Model D | | Extended Model D | |
|---|---|---|---|---|
| | Estimate | Std. Error | Estimate | Std. Error |
| $a_{11}$ | 0.7079 | 0.1211 | 0.5135 | 0.3421 |
| $a_{21}$ | - | - | - | - |
| $a_{12}$ | - | - | - | - |
| $a_{22}$ | 0.8730 | 0.1142 | 0.7649 | 0.1381 |
| $b_{11}$ | 0.1924 | 0.1674 | 0.0049 | 0.5434 |
| $b_{22}$ | 0.0072 | 0.8430 | 0.0006 | 0.0690 |
| $\rho$ | -0.7995 | 0.0029 | -0.4523 | 0.0882 |
| $\overline{\mu}$ | 0.0131 | 0.1065 | 0.0131 | 0.0021 |
| $\overline{\sigma}$ | 0.0857 | 0.5172 | 0.0857 | 0.0009 |
| $c_{11}$ | - | - | 7.7812 | 2.6462 |
| $c_{12}$ | - | - | 1.0911 | 0.8892 |
| $c_{13}$ | - | - | -38.8899 | 1.5632 |
| $c_{14}$ | - | - | 0.4021 | 0.7437 |
| $c_{15}$ | - | - | -39.1056 | 0.1624 |
| $c_{21}$ | - | - | -1.3562 | 1.7051 |
| $c_{22}$ | - | - | -0.1460 | 0.0644 |
| $c_{23}$ | - | - | -0.1245 | 5.6614 |
| $c_{24}$ | - | - | -5.4407 | 1.4016 |
| $c_{25}$ | - | - | 10.7989 | 0.4681 |
| $\mathcal{L}$ | 244.69 | | 263.46 | |

**Estimation results.** This table presents the QML estimates of the model of the form:

$$y_t = \mu(S_{t-1}) + \lambda_(S_{t-1})\varepsilon_t,$$

and

$$S_t = Cx_t + AS_{t-1} + \underline{\eta_t} \text{ with } \underline{\eta_t} \sim \mathcal{N}(0, \Sigma),$$

where

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} \\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} \end{bmatrix}, A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \Sigma = \begin{bmatrix} b_{11} & \rho\sqrt{b_{11}b_{22}} \\ \rho\sqrt{b_{11}b_{22}} & b_{22} \end{bmatrix},$$

$$\mu(S_t) = \overline{\mu}\exp(S_{1t}) \text{ and } \sigma(S_t) = \overline{\sigma}\exp(S_{2t}).$$

The vector of conditioning variables $x_t$ contains the de-meaned consumption, wealth, income ratio ($cay$); log dividend-price ratio ($d-p$); the relative bill rate ($RREL$); the term spread, the difference between the ten-year Treasury bond yield and the three-month Treasury bond yield ($TRM$); and the Baa-Aaa default spread ($DEF$). Heteroscedasticity-consistent standard errors are reported. The estimates are for quarterly returns on the value-weighted CRSP index in excess of the three-month Treasury bill from the second quarter of 1953 to the fourth quarter of 2011.

**Table 3.11.** Summary Statistics of Sharpe Ratio Estimates

| | **Mean** | **Std. Dev.** | **Min.** | **Max.** | **A.C.(1)** |
|---|---|---|---|---|---|
| BK | 0.25 | 0.05 | 0.14 | 0.41 | 0.61 |
| BK (Unconstrained) | 0.26 | 0.10 | -0.05 | 0.49 | 0.71 |
| BK (Extended) | 0.31 | 0.25 | 0.07 | 1.58 | 0.88 |
| OLS Methods | 0.30 | 0.42 | -0.61 | 1.84 | 0.81 |

**Moment comparison.** This table reports descriptive statistics of the estimates of Sharpe ratios based on quarterly realized log-returns for the CRSP VW index for the first quarter of 1953 to the last quarter of 2011. The first, second and third sets of Sharpe ratio estimates are based on the reduced form model by Brandt and Kang (2004) (BK) in which the conditional mean and volatility of stock returns are treated as latent variables. The first representation guarantees positive values for the conditional mean and volatility, while the second representation guarantees a positive volatility only. The third representation is an extended version in which the conditional moments are positive functions of exogenous predictors. Finally, the last set of Sharpe ratio estimates is based on the conditional moments estimated from OLS regressions of log-returns on lagged explanatory variables.

# Chapter 4

# Conclusion

In this thesis, I investigate filtering methods and some applications of asset pricing. Chapter 2 extends the nonlinear filtering literature by proposing a new filtering method based on efficient Taylor approximations. The filter can be applied to estimate latent variables and for parameter inference. I find that the filtering methods that are based on Taylor approximations generate state estimates that are as accurate as the estimates obtained with Monte Carlo filters, while being computationally more efficient.

The filter can be applied in finance and economics where stochastic volatility has become the standard paradigm. I test the filter in three models, namely, in the standard stochastic volatility model (Ghysels, Harvey, and Renault, 1996), in a model of risk and returns (Brandt and Kang, 2004) and in a dynamic stochastic general equilibrium model (Flury and Shephard, 2011). In all these applications, I find that the filter generates accurate state estimates at least five times faster than standard particle filters. I also show how these filters, along with perturbation methods, can be applied to estimate dynamic stochastic general equilibrium models. Finally, by conducting a set of robustness checks I also find that the Taylor series filter is accurate in highly nonlinear and highly dimensional state-space models.

Filtering methods can be naturally applied in finance, where investors and managers take important decisions based on noisy information. These findings are significant, since a more efficient use of filtering methods in finance can help agents to take more informed decisions as well as enable them to identify underlying risks in the economy.

Chapter 3 investigates the dynamic behavior of the market Sharpe ratio. I have examined estimates of the Sharpe ratio volatility and question whether they are biased due to limitations of the empirical methodology used to estimate them. I show that measurement error in estimated Sharpe ratios explains the high time-series variation documented in the empirical literature (Lettau and Ludvigson, 2010). I also show that filtering methods are a better approach to assess the Sharpe ratio's time variation.

My findings have significant implications for portfolio allocation, especially for investors who face proportional costs by trading the mean-variance optimal portfolio. As the optimal weight is proportional to the market Sharpe ratio, the percentage of wealth traded in each period will depend upon the volatility of the market Sharpe ratio. It is clear that upward-biased estimates of the Sharpe ratio volatility would imply excessive portfolio rebalancing, and therefore more transaction costs.

## 4.1 Limitations

Both chapters have room for improvement. In the first chapter, we learned that although the nonlinear filters with Taylor approximations provide accurate results for a number of problems, some care should be taken in their modeling and implementation. For the modeling part, a significant amount of work has to be done to select observation equations that generate accurate state and parameter estimates. Moreover, the filter may diverge when the functions that define the observation equation or transition equation of the state-space models are not differentiable or the Taylor series approximations are not uniformly convergent.

Likewise, in the essay on the volatility of the market Sharpe ratio, filtering is just one technique that provides a correct estimate of the volatility of the market Sharpe ratio. There may be other techniques that could provide accurate measures of the market Sharpe ratio variability, such as the mixed data sampling (MIDAS) regression models (Ghysels, Santa-Clara, and Valkanov, 2005, 2006) and related econometric methods that are based on data sampled at different frequencies.

Lastly, the CRSP index is a proxy for a market portfolio that is unobservable. The measurement error generated by this proxy generates some additional variation that I am unable to explicitly account for in the current setup. By identifying this source of variation, we would be able to know to what extent the Sharpe ratio variability is generated by the estimation methodology, the measurement error of the market portfolio or any source of risk.

## 4.2 Future Work

The filtering techniques developed in Chapter 2 have the flexibility to extend to non-differentiable functions or non-Gaussian errors. This can be done by using a basis of orthogonal polynomials (such as the Chebyshev, Hermite, Legendre or Laguerre polynomials) to approximate nonlinear functions, instead of using the Taylor series. As a result, more general state-space models can be analyzed through this filtering setup.

In addition to the study of non-differentiable functions, there are a number of extensions that I plan to follow based on this setup, such as optimal selection of observation equations; the addition of extra centers of expansions to conduct better state and parameter estimation; the use of Bayesian techniques for parameter estimation and finally, the application of these techniques in Markov switching models.

As for Chapter 3, a number of extensions can be done. The most natural application consists of formally studying the implications of filtering in optimal portfolio allocation and predictability, such as in Section 3.6. A large body of empirical work has found evidence of predictability of returns in equity and across other financial markets. [65] Such predictability has important implications for the asset allocation of investors. The empirical evidence suggests that long-term investors should take into account the predictability in strategic asset allocation (Campbell and Viceira, 2002). In particular, one strand of the literature advocates, in principle, the countercyclical market timing of the portfolio equity share, using predictive variables for expected returns. In the standard portfolio allocation setup, rebalancing to a fixed strategic equity share ignores the time variation in the equity risk premium. If such variation is real and persists in a similar way to the historical asset return data, a fixed strategic equity weight would not be optimal. The rebalancing of the equity share should therefore be modified so as to incorporate predictability within the asset allocation setup. Filtering techniques are the natural approach to incorporate such predictability.

Another natural extension of Chapter 3 is based on the analysis of the cross-section of the equity Sharpe ratios which can vary significantly by the characteristics of the firm or the portfolio. We learn from Chapter 3 that the Sharpe ratio of any asset is unobservable. However, filtering techniques are the natural econometric techniques to be applied for inference purposes. Another important extension is related to the data-based performance measures for asset pricing models. Backus, Chernov, and Zin (2012) introduce entropy as a measure for capturing the dispersion and time-series dependence of a model-implied pricing kernel. This measure is linearly related to the volatility of the market Sharpe ratio. In this extension I intend to empirically quantify the entropy of asset pricing models via filtering methods and to relate it to model-implied measures, such as the market Sharpe ratio and its volatility.

Even 40 years after the discovery of filtering techniques, we still have the capacity for theoretical and empirical applications. Given that, in most cases, investors and econometricians only have partial information, I posit that filtering techniques are a powerful tool that should be used to analyze dynamic models.

[65] More recently, the evidence on predictability has grown and become inconclusive (Boudoukh, Richardson, and Whitelaw, 2008; Campbell and Shiller, 1988; Fama and French, 1989; Goyal and Welch, 2008; Hodrick, 1992; Keim and Stambaugh, 1986; Lewellen, 2004; Stambaugh, 1999; Valkanov, 2003).

# Bibliography


Abel, A. B., 1990. Asset Prices under Habit Formation and Catching up with the Joneses. The American Economic Review 80(2), pp. 38–42. → pages 83, 104

An, S., F. Schorfheide, 2007. Bayesian Analysis of DSGE Models. Econometric Reviews 26(2-4), 113–172. → pages 32

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. → pages 4, 20, 86

Andersen, T. G., T. Bollerslev, F. X. Diebold, P. Labys, 2003. Modeling and Forecasting Realized Volatility.. Econometrica 71(2), 579. → pages 4, 20, 86

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. → pages 4, 19, 27, 86

Anderson, B., J. B. Moore, 1979. Optimal Filtering vol. 1, . → pages 8

Ang, A., G. Bekaert, 2007. Stock Return Predictability: Is it There?. The Review of Financial Studies 20(3), pp. 651–707. → pages 77

Ang, A., J. Liu, 2007. Risk, return, and dividends. Journal of Financial Economics 85(1), 1 – 38. → pages 85

Ang, A., M. Piazzesi, 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50(4), 745 – 787. → pages 72

Asai, M., M. McAleer, J. Yu, 2006. Multivariate Stochastic Volatility: a Review. Econometric Reviews 25(2-3), 145–175. → pages 41

Back, K., 2010. Asset Pricing and Portfolio Choice Theory. Oxford University Press, . → pages 141

Backus, D., M. Chernov, S. E. Zin, 2012. Sources of Entropy in Representative Agent Models. Working Paper New York University. → pages 84, 114

Bansal, R., D. Kiku, A. Yaron, 2012a. An Empirical Evaluation of the Long-Run Risks Model for Asset Prices. Critical Finance Review 1, 183–221. → pages 73, 143, 147

——— , 2012b. Risks For the Long Run: Estimation with Time Aggregation. Working Paper Duke University. → pages 75

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. → pages 4, 26, 68, 69, 71, 73, 83, 95, 96, 97, 101, 102, 103, 104

Barillas, F., L. P. Hansen, T. J. Sargent, 2009. Doubts or variability?. Journal of Economic Theory 144(6), 2388 – 2418. → pages 25

Barro, R. J., 2006. Rare Disasters and Asset Markets in the Twentieth Century. The Quarterly Journal of Economics 121(3), pp. 823–866. → pages 83, 104

——— , 2009. Rare Disasters, Asset Prices, and Welfare Costs. The American Economic Review 99(1), pp. 243–264. → pages 83, 104

Basak, S., D. Cuoco, 1998. An Equilibrium Model with Restricted Stock Market Participation. The Review of Financial Studies 11(2), pp. 309–341. → pages 83

Beeler, J., J. Y. Campbell, 2012. The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment. Critical Finance Review 1, 141–182. → pages 76, 147

Bekaert, G., C. R. Harvey, 1995. Time-Varying World Market Integration. The Journal of Finance 50(2), pp. 403–444. → pages 86

Bhamra, H. S., R. Uppal, 2010. Asset Prices with Heterogeneity in Preferences and Beliefs. Working Paper Edhec Business School and University of British Columbia. → pages 83

Bidder, R., M. Smith, 2011. Doubts and Variability. New York University Working Paper. → pages 4, 25, 26

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, S. Terry, 2012. Really Uncertain Business Cycles. Working Paper: Stanford University. → pages 25

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. → pages 4

Bollerslev, T., R. Engle, J. Wooldridge, 1988. A capital asset pricing model with time-varying covariances. The Journal of Political Economy pp. 116–131. → pages 77

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. → pages 4, 8, 18, 19

Boudoukh, J., M. Richardson, R. F. Whitelaw, 2008. The Myth of Long-Horizon Predictability. Review of Financial Studies 21(4), 1577–1605. → pages 114

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. → pages 4, 19, 26, 29, 30, 31, 32, 47, 48, 49, 50, 70, 71, 72, 74, 84, 85, 87, 88, 89, 90, 91, 108, 111, 112, 143

Breeden, D. T., 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. Journal of Financial Economics 7(3), 265 – 296. → pages 83, 104

Breen, W., L. R. Glosten, R. Jagannathan, 1989. Economic Significance of Predictable Variations in Stock Index Returns. The Journal of Finance 44(5), pp. 1177–1189. → pages 77

Broto, C., E. Ruiz, 2004. Estimation Methods for Stochastic Volatility Models: a Survey.. Journal of Economic Surveys 18(5), 613 – 649. → pages 4, 20, 22

Calvet, L., A. J. Fisher, L. Wu, 2013. Staying on Top of the Curve: A Cascade Model of Term Structure Dynamics. Working Paper: University of British Columbia. → pages 18, 82

Calvet, L. E., A. J. Fisher, 2007. Multifrequency News and Stock Returns. Journal of Financial Economics 86(1), 178 – 212. → pages 83, 104

Campbell, J., R. Shiller, 1988. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1(3), 195–228. → pages 73, 74, 114, 143

Campbell, J., A. Sunderam, L. Viceira, 2011. Inflation bets or deflation hedges? the changing risks of nominal bonds. Working Paper: Harvard University. → pages 18

——— , 2012. Inflation Bets or Deflation Hedges? The Changing Risks of Nominal Bonds. Working Paper Harvard University. → pages 82

Campbell, J., L. Viceira, 2002. Strategic Asset Allocation: Portfolio Choice for Long-Term Investors. Oxford University Press, USA, . → pages 93, 114, 143

Campbell, J. Y., 1987. Stock returns and the term structure. Journal of Financial Economics 18(2), 373 – 399. → pages 77, 85, 87

——— , 1991. A Variance Decomposition for Stock Returns. The Economic Journal 101(405), pp. 157–179. → pages 77

Campbell, J. Y., J. H. Cochrane, 1999. By Force of Habit: A Consumption Based Explanation of Aggregate Stock Market Behavior. Journal of Political Economy 107(2), pp. 205–251. → pages 68, 71, 75, 83, 84, 88, 98, 104, 153, 154

Campbell, J. Y., L. Hentschel, 1992a. No news is good news: An asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31(3), 281 – 318. → pages 32

——— , 1992b. No news is good news: An asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31(3), 281 – 318. → pages 74, 87

Casella, G., C. P. Robert, 1996. Rao-Blackwellisation of Sampling Schemes. Biometrika 83(1), 81–94. → pages 17

——— , 2004. Monte Carlo Statistical Methods vol. 319. Citeseer, . → pages 137, 138

Chan, Y. L., L. Kogan, 2002. Catching Up with the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices.. Journal of Political Economy 110(6), 1255 – 1285. → pages 83, 104

Chib, S., F. Nardari, N. Shephard, 2006. Analysis of High Dimensional Multivariate Stochastic Volatility Models. Journal of Econometrics 134(2), 341–371. → pages 39, 42, 55

Chib, S., Y. Omori, M. Asai, 2009. Multivariate Stochastic Volatility. pp. 365–400. → pages 39, 41

Chien, Y., H. Cole, H. Lustig, 2012. Is the Volatility of the Market Price of Risk Due to Intermittent Portfolio Rebalancing?. American Economic Review 102(6), 2859–96. → pages 68, 83, 104

Christoffersen, P., K. Jacobs, L. Karoui, K. Mimouni, 2012. Nonlinear Kalman Filtering in Affine Term Structure Models. Working Paper: University of Toronto. → pages 18, 133

Clark, T., 2009. Is the Great Moderation over? An Empirical Analysis. Economic Review 4, 5–42. → pages 25

Conrad, J., G. Kaul, 1988. Time-Variation in Expected Returns. The Journal of Business 61(4), pp. 409–425. → pages 72

Constantinides, G. M., 1990. Habit Formation: A Resolution of the Equity Premium Puzzle. Journal of Political Economy pp. 519 – 543. → pages 71, 83, 88, 104

Constantinides, G. M., D. Duffie, 1996. Asset Pricing with Heterogeneous Consumers. Journal of Political Economy 104(2), pp. 219–240. → pages 83

Crisan, D., B. Rozovskii, 2011. The Oxford handbook of Nonlinear Filtering. Oxford University Press, . → pages 7, 17, 69, 81

Croushore, D., T. Stark, 2001. A real-time data set for macroeconomists. Journal of Econometrics 105(1), 111 – 130. → pages 26

Dangl, T., M. Halling, 2012. Predictive regressions with time-varying coefficients. Journal of Financial Economics 106(1), 157 – 181. → pages 72

De Miguel, V., L. Garlappi, R. Uppal, 2009. Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?. The Review of Financial Studies 22(5), pp. 1915–1953. → pages 93

DeJong, D. N., C. Dave, 2011. Structural Macroeconometrics. Princeton University Press, . → pages 34

Doucet, A., N. de Freitas, N. Gordon, A. Smith, 2001. Sequential Monte Carlo Methods in Practice. Springer, . → pages 6, 69, 81

Doucet, A., N. De Freitas, K. Murphy, S. Russell, 2000. Rao-Blackwellised particle filtering for dynamic Bayesian networks. Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence pp. 176–183. → pages 4, 6, 17

Duffee, G. R., 1996. Idiosyncratic Variation of Treasury Bill Yields. The Journal of Finance 51(2), pp. 527–551. → pages 87

Einicke, G. A., 2012. Smoothing, Filtering and Prediction: Estimating the Past, Present and Future. Rijeka, Croatia: Intech. ISBN, . → pages

Epstein, L. G., S. E. Zin, 1989. Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. Econometrica 57(4), pp. 937–969. → pages 83

Fama, E. F., K. R. French, 1988a. Dividend yields and expected stock returns. Journal of Financial Economics 22(1), 3 – 25. → pages 77

——— , 1988b. Permanent and Temporary Components of Stock Prices. Journal of Political Economy 96(2), pp. 246–273. → pages 27, 86

——— , 1989. Business conditions and Expected Returns on Stocks and Bonds. Journal of Financial Economics 25(1), 23–49. → pages 114

Fama, E. F., G. Schwert, 1977. Asset returns and inflation. Journal of Financial Economics 5(2), 115 – 146. → pages 77

Fernández-Villaverde, J., J. F. Rubio-Ramírez, 2007. Estimating Macroeconomic Models: A Likelihood Approach.. Review of Economic Studies 74(4), 1059 – 1087. → pages 3, 5, 25, 33

Fernández-Villaverde, J., J. F. Rubio-Ramírez, M. S. Santos, 2006. Convergence Properties of the Likelihood of Computed Dynamic Models. Econometrica 74(1), 93–119. → pages 3, 33

Ferson, W. E., S. Sarkissian, T. T. Simin, 2003. Spurious Regressions in Financial Economics?. The Journal of Finance 58(4), pp. 1393–1413. → pages 69

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. → pages 32, 35, 38, 39, 112

French, K. R., G. Schwert, R. F. Stambaugh, 1987. Expected stock returns and volatility. Journal of Financial Economics 19(1), 3 – 29. → pages 32, 77

Gallant, A., H. White, 1988. A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Basil Blackwell New York, . → pages 18, 19, 139

Garlappi, L., G. Skoulakis, 2010. Solving Consumption and Portfolio Choice Problems: The State Variable Decomposition Method. Review of Financial Studies 23(9), 3346–3400. → pages 5, 10, 13

——— , 2011. Taylor series approximations to expected utility and optimal portfolio choice. Mathematics and Financial Economics 5, 121–156. → pages 11, 13

Gârleanu, N., S. Panageas, 2011. Young, Old, Conservative and Bold: The Implications of Heterogeneity and Finite Lives for Asset Pricing. Working Paper University of Chicago. → pages 83

Ghysels, E., A. C. Harvey, E. Renault, 1996. Stochastic volatility vol. 14. Elsevier North-Holland, . → pages 20, 112

Ghysels, E., P. Santa-Clara, R. Valkanov, 2005. There is a Risk-Return Trade-Off After All. Journal of Financial Economics 76(3), 509–548. → pages 85, 113

——— , 2006. Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies. Journal of Econometrics 131, 59 – 95. → pages 77, 113

Glosten, L. R., R. Jagannathan, D. E. Runkle, 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. The Journal of Finance 48(5), pp. 1779–1801. → pages 74, 77, 85

Gomes, F., A. Michaelides, 2008. Asset Pricing with Limited Risk Sharing and Heterogeneous Agents. Review of Financial Studies 21(1), 415 – 448. → pages 83

Gordon, N. J., D. J. Salmond, A. F. Smith, 1993. Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation. IEE Proceedings F (Radar and Signal Processing) 140(2), 107–113. → pages 6, 39, 41, 54

Goyal, A., I. Welch, 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. The Review of Financial Studies 21(4), pp. 1455–1508. → pages 77, 87, 114

Guvenen, F., 2009. A Parsimonious Macroeconomic Model for Asset Pricing. Econometrica 77(6), pp. 1711–1750. → pages 83

Hamilton, J. D., 1994. Time Series Analysis. Princeton University Press, . → pages 6, 27, 32, 69, 85

Hansen, L. P., R. Jagannathan, 1991. Implications of Security Market Data for Models of Dynamic Economies. Journal of Political Economy 99(2), pp. 225–262. → pages 142

Hansen, L. P., T. J. Sargent, 1983. Aggregation Over Time and the Inverse Optimal Predictor Problem for Adaptive Expectations in Continuous Time. International Economic Review 24(1), pp. 1–20. → pages 75

Harrison, J., D. Kreps, 1979. Martingales and arbitrage in multiperiod securities markets. Journal of Economic Theory 20(3), 381–408. → pages 72, 141

Harvey, A., E. Ruiz, N. Shephard, 1994. Multivariate Stochastic Variance Models. The Review of Economic Studies 61(2), pp. 247–264. → pages 42

Harvey, A. C., N. Shephard, 1996. Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns. Journal of Business & Economic Statistics 14(4), pp. 429–434. → pages 22

Harvey, C. R., 2001. The Specification of Conditional Expectations. Journal of Empirical Finance 8(5), 573 – 637. → pages 85

Heaton, J., 1995. An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications. Econometrica 63(3), pp. 681–717. → pages 75

Hodrick, R. J., 1992. Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement. The Review of Financial Studies 5(3), pp. 357–386. → pages 77, 114

Ito, K., K. Xiong, 2000. Gaussian Filters for Nonlinear Filtering Problems. Automatic Control, IEEE Transactions on 45(5), 910 –927. → pages 4, 8, 23

Jacquier, E., M. Johannes, N. Polson, 2007. MCMC maximum likelihood for latent state models. Journal of Econometrics 137(2), 615 – 640. → pages 27, 86

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. → pages 27

Jazwinski, A. H., 1970. Stochastic Processes and Filtering Theory. Academic Press, . → pages 5

Jermann, U. J., 2010. The equity premium implied by production. Journal of Financial Economics 98(2), 279 – 296. → pages 83, 104

Jiming, J., 2010. Large Sample Techniques for Statistics. Springer, . → pages 11

Johannes, M., N. Polson, 2009. Particle Filtering. pp. 1015–1029. → pages 136

Johnson, T., J. Lee, 2012. Systematic Volatility of Unpriced Earnings Shocks. University of Illinois, Urbana-Champaign Working Paper. → pages 76

Judd, K. L., 1998. Numerical Methods in Economics. MIT Press (Cambridge, Mass.), . → pages 5, 34

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. → pages 5, 8, 133

——— , 2004. Unscented Filtering and Nonlinear Estimation. Proceedings of the IEEE 92(3), 401 – 422. → pages 133

Jungbacker, B., S. J. Koopman, 2006. Monte Carlo Likelihood Estimation for Three Multivariate Stochastic Volatility Models. Econometric Reviews 25(2-3), 385–408. → pages 43

Justiniano, A., G. E. Primiceri, 2008. The Time-Varying Volatility of Macroeconomic Fluctuations. American Economic Review 98(3), 604–41. → pages 25

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. → pages 3, 7, 10, 131

Kandel, S., R. F. Stambaugh, 1990. Expectations and Volatility of Consumption and Asset Returns. The Review of Financial Studies 3(2). → pages 77

Keim, D. B., R. F. Stambaugh, 1986. Predicting returns in the stock and bond markets. Journal of Financial Economics 17(2), 357 – 390. → pages 114

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, . → pages 6, 69

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. → pages 27, 86

Koopman, S. J., G. Sandmann, 1998. Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood. Journal of Econometrics 87(2), 271 – 301. → pages 22

Lamoureux, C., G. Zhou, 1996. Temporary components of stock returns: What do the data tell us?. Review of Financial Studies 9(4), 1033–1059. → pages 27, 72, 86, 90

Lettau, M., S. Ludvigson, 2001a. Consumption, Aggregate Wealth, and Expected Stock Returns. The Journal of Finance 56(3), 815–849. → pages 77

——— , 2001b. Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time Varying. Journal of Political Economy 109(6), pp. 1238–1287. → pages 77

——— , 2010. Measuring and Modeling Variation in the Risk-Return Trade-Off. Handbook of Financial Econometrics. → pages 68, 71, 77, 85, 87, 88, 89, 112

Lettau, M., H. Uhlig, 2002. The Sharpe Ratio and preferences: A Parametric Approach. Macroeconomic Dynamics 6(02), 242–265. → pages 142

Lettau, M., S. Van Nieuwerburgh, 2008. Reconciling the Return Predictability Evidence. Review of Financial Studies 21(4), 1607–1652. → pages 69

Lewellen, J., 1999. The time-series relations among expected return, risk, and book-to-market. Journal of Financial Economics 54(1), 5 – 43. → pages 77

——— , 2004. Predicting returns with financial ratios. Journal of Financial Economics 74(2), 209 – 235. → pages 114

Lintner, J., 1965. Security Prices, Risk, and Maximal Gains From Diversification. The Journal of Finance 20(4), pp. 587–615. → pages 83

Lucas, R., 1978. Asset Prices in an Exchange Economy. Econometrica pp. 1429 – 1445. → pages 83, 104

Ludvigson, S., 2012. Advances in Consumption-Based Asset Pricing. Working Paper: New York University. → pages 4, 26

Ludvigson, S. C., S. Ng, 2007. The Empirical Risk-Return Relation: A Factor Analysis Approach. Journal of Financial Economics 83(1), 171 – 222. → pages 71, 77, 85

Lundblad, C., 2007. The Risk-return Trade-off in the long run: 1836-2003. Journal of Financial Economics 85(1), 123 – 150. → pages 32, 85

Lustig, H. N., A. Verdelhan, 2012. Business Cycle Variation in the Risk-Return Trade-Off. Journal of Monetary Economics (Forthcoming). → pages 71, 88

Martin, I., 2012. Simple Variance Swaps. Working Paper Stanford University. → pages 84

Maz'ya, V., G. Schmidt, 1996. On Approximate Approximations Using Gaussian Kernels. IMA Journal of Numerical Analysis 16(1), 13–29. → pages 4

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). → pages 25

Mehra, R., E. C. Prescott, 1985. The equity premium: A puzzle. Journal of Monetary Economics 15(2), 145 – 161. → pages 83, 104

Melsa, J. L., A. P. Sage, 1971. Estimation Theory with Applications to Communications and Control. McGraw-Hill Book Company, . → pages

Menzly, L., T. Santos, P. Veronesi, 2004. Understanding Predictability. Journal of Political Economy 112(1), pp. 1–47. → pages 104

Merton, R. C., 1969. Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case. The Review of Economics and Statistics 51(3), pp. 247–257. → pages 71

Merton, R. C., 1971. Optimum consumption and portfolio rules in a continuous-time model. Journal of Economic Theory 3(4), 373 – 413. → pages 71

Merton, R. C., 1973. An Intertemporal Capital Asset Pricing Model. Econometrica 41(5), pp. 867–887. → pages 85

——— , 1980. On estimating the expected return on the market: An exploratory investigation. Journal of Financial Economics 8(4), 323 – 361. → pages 86

Newey, W., K. West, 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55(3), 703–708. → pages 139

Newey, W. K., K. D. West, 1994. Automatic Lag Selection in Covariance Matrix Estimation. The Review of Economic Studies 61(4), 631–653. → pages 139

Nielsen, L. T., M. Vassalou, 2004. Sharpe Ratios and Alphas in Continuous Time. The Journal of Financial and Quantitative Analysis 39(1), pp. 103–114. → pages 143

Papanikolaou, D., 2011. Investment Shocks and Asset Prices. Journal of Political Economy 119(4), pp. 639–685. → pages 83, 104

Pastor, L., M. Sinha, B. Swaminathan, 2008. Estimating the Intertemporal Risk-Return Tradeoff Using the Implied Cost of Capital. The Journal of Finance 63(6), pp. 2859–2897. → pages 85

Pitt, M. K., N. Shephard, 1999. Filtering via Simulation: Auxiliary Particle Filters.. Journal of the American Statistical Association 94(446), 590 – 599. → pages 6

Pástor, L., R. F. Stambaugh, 2009. Predictive Systems: Living with Imperfect Predictors.. Journal of Finance 64(4), 1583 – 1628. → pages 72

Rietz, T. A., 1988. The equity risk premium a solution. Journal of Monetary Economics 22(1), 117 – 131. → pages 83, 104

Romero, A., 2012. Filtering via Taylor Series Approximations. Working Paper: University of British Columbia. → pages

Ruiz, E., 1994. Quasi-maximum likelihood estimation of stochastic volatility models.. Journal of Econometrics 63(1), 289 – 306. → pages 22

Rytchkov, O., 2012. Filtering out Expected Dividends and Expected Returns. Quarterly Journal of Finance (Forthcoming). → pages 39, 69, 72

Savits, T. H., 2006. Some statistical applications of Faa di Bruno. Journal of Multivariate Analysis 97(10), 2131 – 2140. → pages 5, 6, 10, 11, 12, 13, 126, 127

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. → pages 5, 19, 35, 38, 67

Schwert, G. W., 1989. Why Does Stock Market Volatility Change Over Time?. The Journal of Finance 44(5), pp. 1115–1153. → pages 77

Scruggs, J. T., 1998. Resolving the Puzzling Intertemporal Relation between the Market Risk Premium and Conditional Market Variance: A Two-Factor Approach. The Journal of Finance 53(2), pp. 575–603. → pages 85

Sharpe, W. F., 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. The Journal of Finance 19(3), pp. 425–442. → pages 83

Shephard, N., 1996. Statistical Aspects of ARCH and Stochastic Volatility. Monographs on Statistics and Applied Probability 65, 1–68. → pages 42

Shephard, N. E., 2005. Stochastic Volatility: Selected Readings. Oxford University Press, USA, . → pages 20

Simon, D., 2006. Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches. Wiley-Interscience, . → pages

Stambaugh, R. F., 1999. Predictive regressions. Journal of Financial Economics 54(3), 375–421. → pages 114

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

Stock, J., M. Watson, 2002. Has the Business Cycle Changed and Why?. NBER Working Papers. → pages 25

Tallarini, T. D. J., 2000. Risk-sensitive real business cycles. Journal of Monetary Economics 45(3), 507 – 532. → pages 25

Tang, Y., R. F. Whitelaw, 2011. Time-Varying Sharpe Ratios and Market Timing. Quarterly Journal of Finance 1(3), 465 – 493. → pages 71, 77, 88

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

Taylor, S., 2008. Modelling Financial Time Series. World Scientific, . → pages 20

Valkanov, R., 2003. Long-horizon regressions: theoretical results and applications. Journal of Financial Economics 68(2), 201 – 232. → pages 114

van Binsbergen, J., J. Fernandez-Villaverde, R. Koijen, J. F. Rubio-Ramirez, 2012. The Term Structure of Interest Rates in a DSGE Model with Recursive Preferences. Journal of Monetary Economics: Forthcoming. → pages 81

van Binsbergen, J., R. Koijen, 2011. Likelihood-based Estimation of Exactly-Solved Present-Value Models. Working Paper: University of Chicago. → pages 18, 82, 133

van Binsbergen, J. H., R. S. J. Koijen, 2010. Predictive Regressions: A Present-Value Approach.. Journal of Finance 65(4), 1439 – 1471. → pages 72

Van Der Merwe, R., A. Doucet, N. De Freitas, E. Wan, 2001. The Unscented Particle Filter. Advances in Neural Information Processing Systems pp. 584–590. → pages 6

Vuolteenaho, T., 2000. Understanding the Aggregate Book-to-Market Ratio. Unpublished paper, Harvard University. → pages 77

Wachter, J. A., 2005. Solving models with external habit. Finance Research Letters 2(4), 210 – 226. → pages 153, 154, 155

——— , 2012. Can time-varying risk of rare disasters explain aggregate stock market volatility?. Journal of Finance, Forthcoming. → pages 83, 104

Weil, P., 1989. The equity premium puzzle and the risk-free rate puzzle. Journal of Monetary Economics 24(3), 401 – 421. → pages 83, 104

White, H., 1982. Maximum Likelihood Estimation of Misspecified Models. Econometrica 50(1), pp. 1–25. → pages 4, 8, 18

Whitelaw, R., 2000. Stock market risk and return: An equilibrium approach. Review of Financial Studies 13(3), 521 – 547. → pages 85

Whitelaw, R. F., 1994. Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns. The Journal of Finance 49(2), pp. 515–541. → pages 77, 88

Winschel, V., M. Krätzig, 2010. Solving, Estimating, and Selecting Nonlinear Dynamic Models Without the Curse of Dimensionality. Econometrica 78(2), 803–821. → pages 5, 8

# Appendix A

# Appendix to Chapter 2

## A.1 Efficient Calculation of Derivatives of Composite Functions

The efficient calculation of partial derivatives described in Chapter 2 relies on the Taylor expansion of a function of the form $f(\mathbf{x}) = h(g(\mathbf{x}))$, where $h : \mathbb{R} \longrightarrow \mathbb{R}$, $g : \mathbb{R}^N \longrightarrow \mathbb{R}$, and $\mathbf{x} = (x_1, x_2, ..., x_N)$ denotes an $N$-dimensional vector.[66] The generic $M$-th order Taylor expansion of $f$ centered at a constant point $\mathbf{0}_N$ is defined as

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

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 $f(\mathbf{x})$ evaluated at $\mathbf{0}_N$; i.e.,

$$f_{\mathbf{q}}(\mathbf{0}_N) = \frac{\partial^{q_1+...+q_N} f}{\partial x_1^{q_1} ... \partial x_1^{q_N}}(\mathbf{0}_N). \tag{A.2}$$

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 $q_n \in \mathbb{N}_0, n = 1, \ldots, N$. We write $\ell \leq \mathbf{q}$ if $\ell_n \leq q_n$, for $n = 1, \ldots, N$, and denote

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

Let $g_{\mathbf{q}}(\mathbf{x})$ denote the partial derivative of order $\mathbf{q}$ of the function $g(\mathbf{x})$, 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

[66] For simplicity I consider the case in which $f(\mathbf{x})$ 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.

composite function $f(x) = h(g(\mathbf{x}))$; i.e., $f_{\mathbf{q}}(\mathbf{x})$ can be expressed as

$$f_{\mathbf{q}}(\mathbf{x}) = \sum_{n=1}^{|\mathbf{q}|} h_n(g(\mathbf{x}))\, \alpha_{\mathbf{q},n}(\mathbf{x}), \tag{A.3}$$

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

**Theorem A.1.1** *For* $\mathbf{q} \geq \mathbf{0}_N$, $1 \leq j \leq N$, *and* $1 \leq n \leq |\mathbf{q}| + 1$, *we have*

$$\alpha_{\mathbf{q}+\mathbf{e_j},n}(\mathbf{x}) = \sum_{\left\{\ell \in \mathbb{N}_0^N:\ \mathbf{0}_N \leq \ell \leq \mathbf{q};|\ell| \geq n-1\right\}} \binom{\mathbf{q}}{\ell} g_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\, \alpha_{\ell,n-1}(\mathbf{x}), \tag{A.4}$$

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

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

***Proof*** *See Savits (2006).*

## A.2 Proofs

**Lemma A.2.1** *Let $Z \sim \mathcal{N}(\mu, \sigma^2)$ and let $f$, any continuously differentiable function such that $f'$ exists almost everywhere and $\mathbb{E}|f'(Z)| < \infty$; then*

$$cov(Z, f(Z)) = \mathbb{E}[(Z-\mu) f(Z)] = \sigma^2 \mathbb{E}[f'(Z)].$$

**Proof**

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

∎

**Proof of Lemma 2.3.5**. Let $\mathbf{Z} = (Z_1, ..., Z_N)$, where $Z_i$ are i.i.d. $\mathcal{N}(0,1)$ random variables. From the previous lemma, we know that for any $g(X)$, differentiable almost everywhere, $cov[Z_i, g(\mathbf{Z})] = \mathbb{E}[\partial g / \partial z_i]$, therefore

$$cov[\mathbf{Z}, g(\mathbf{Z})] = \mathbb{E}[\nabla g(\mathbf{Z})]. \tag{A.5}$$

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 side of (2.18) is

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

∎

**Proof of Proposition 2.3.3**. Let $h(y) = y^2$; then $h_0(y) = y^2$; $h_1(y) = 2y$ and $h_2(y) = 2$. Now, by applying Theorem A.1.1 to $f(\mathbf{x}) = h(g(\mathbf{x}))$, where $h(\cdot)$ is defined as before, we have

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

Now, from Eq. $(A.4)$ applied to $n = 1, 2$, we have

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

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

Finally, we have

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

as required.

∎

**Proof of Proposition 2.3.4** 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,

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

By applying Lemma to the function $\left(g^1(\mathbf{x})+g^2(\mathbf{x})\right)^2$, we have

$$\begin{aligned}
\left(g^1+g^2\right)^2_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x}) &= \sum_{\left\{\ell\in\mathbb{N}_0^N:\,\mathbf{0}_N\le\ell\le\mathbf{q}\right\}}\binom{\mathbf{q}}{\ell}2\cdot\left(g^1+g^2\right)_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\left(g^1+g^2\right)_{\ell}(\mathbf{x}) \qquad (A.6)\\
&= \sum_{\left\{\ell\in\mathbb{N}_0^N:\,\mathbf{0}_N\le\ell\le\mathbf{q}\right\}}\binom{\mathbf{q}}{\ell}2\cdot\left(g^1_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\,g^1_{\ell}(\mathbf{x})+g^2_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\,g^2_{\ell}(\mathbf{x})\right)\\
&\quad+\sum_{\left\{\ell\in\mathbb{N}_0^N:\,\mathbf{0}_N\le\ell\le\mathbf{q}\right\}}\binom{\mathbf{q}}{\ell}2\cdot\left(g^1_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\,g^2_{\ell}(\mathbf{x})+g^2_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\,g^1_{\ell}(\mathbf{x})\right)\\
&= \left(g^1\right)^2_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x})+\left(g^2\right)^2_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x})\\
&\quad+\sum_{\left\{\ell\in\mathbb{N}_0^N:\,\mathbf{0}_N\le\ell\le\mathbf{q}\right\}}\binom{\mathbf{q}}{\ell}2\cdot\left(g^1_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\,g^2_{\ell}(\mathbf{x})+g^2_{\mathbf{q}+\mathbf{e}_j-\ell}(\mathbf{x})\,g^1_{\ell}(\mathbf{x})\right).
\end{aligned}$$

By subtracting $\left(g^1\right)^2_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x})$ and $\left(g^2\right)^2_{\mathbf{q}+\mathbf{e}_j}(\mathbf{x})$ and dividing by two Eq. $(A.6)$ we get the desired result.

∎

**Proof of Lemma 2.6.1**.

Let $\lambda>\frac{1}{2}$; then

$$\frac{1}{1+x^2}=\frac{1/\lambda}{\frac{1}{\lambda}+\frac{x^2}{\lambda}}=\frac{1/\lambda}{1+\frac{1}{\lambda}+\frac{x^2}{\lambda}-1}=\frac{1}{\lambda}\sum_{j=0}^{\infty}\left(\frac{\lambda-1-x^2}{\lambda}\right)^j \qquad (A.7)$$

and this equality is true, as long as

$$\left|\frac{\lambda-1-x^2}{\lambda}\right|<1$$

or

$$|x|<\sqrt{2\lambda-1}.$$

Finally, multiplying both sides of Eq. $(A.7)$ by $x$ and $x^2$ yields to Eq. (2.66) and Eq. (2.67) respectively. Now, to obtain Eq. (2.68) we take the derivative with respect to $x$ of both sides of Eq. $(A.7)$,

$$\frac{-2x}{(1+x^2)^2}=\frac{1}{\lambda}\sum_{j=0}^{\infty}j\frac{-2x}{\lambda}\left(\frac{\lambda-1-x^2}{\lambda}\right)^{j-1},$$

multiplying both sides of the last equation by $-x/2$, yields to the desired result. ∎

## A.3 Standard Kalman Filter

The state-space representation of a linear model is given by

$$\begin{aligned} y_t &= Hx_t + v_t \\ x_t &= Fx_{t-1} + w_t, \end{aligned} \tag{A.8}$$

where $v_t \sim \mathcal{N}(0, R)$ and $w_t \sim \mathcal{N}(0, Q)$. 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.

In systems like $(A.8)$ where the state variables are normally distributed and the measurement equations are linear, the standard Kalman filter yields 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

$$\begin{aligned} \widehat{x}_{t+1|t} &= F\widehat{x}_{t|t} \\ P_{t+1|t} &= FP_{t|t}F^{\top} + Q. \end{aligned} \tag{A.9}$$

The update rule is given by

$$\begin{aligned} K_{t+1} &= P_{t+1|t}H^{\top}\left[P^{yy}_{t+1|t}\right]^{-1} \\ P^{yy}_{t+1|t} &= HP_{t+1|t}H^{\top} + R \\ \widehat{x}_{t+1|t+1} &= \widehat{x}_{t+1|t} + K_{t+1}\left(y_{t+1} - H\widehat{x}_{t+1|t}\right) \\ P_{t+1|t+1} &= \left(I - K_{t+1}H\right)P_{t+1|t} \end{aligned} \tag{A.10}$$

## A.4 The Extended Kalman Filter

A well known approximation to non linear filtering is the extended Kalman filter (EKF), 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

$$y_{t+1} = h\left(x_{t+1|t}\right) + H_{t+1}\left(x_{t+1} - x_{t+1|t}\right) + v_{t+1}, \tag{A.11}$$

where

$$H_{t+1} = \left.\frac{\partial h}{\partial x_{t+1}}\right|_{x_{t+1}=x_{t+1|t}} \tag{A.12}$$

denotes the Jacobian matrix of the nonlinear function $g$ computed at $x_{t+1|t}$. The transition equation is linearized as in $(A.12)$ and is written as

$$x_{t+1} = g\left(x_{t|t}\right) + G_t\left(x_t - x_{t|t}\right) + \varepsilon_{t+1}, \tag{A.13}$$

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

$$P^{xy}_{t+1|t} = P^{xx}_{t+1|t} H_{t+1}, \tag{A.14}$$

$$P^{yy}_{t+1|t} = H_{t+1} P^{xx}_{t+1|t} H^{\top}_{t+1} + R \tag{A.15}$$

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.

## A.5 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 EKF, 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 EKF as noted by van Binsbergen and Koijen (2011) and Christoffersen, Jacobs, Karoui, and Mimouni (2012).

Let $x$ denote a random vector with mean $\mu_x$ 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 $\mu_x$ 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

$$\mathcal{X}_0 = \mu_x, \tag{A.16}$$

$$\mathcal{X}_i = \mu_x + \left(\sqrt{(N+\xi)P^{xx}}\right)_i, \text{ for } i = 1,\ldots,N \tag{A.17}$$

$$\mathcal{X}_i = \mu_x - \left(\sqrt{(N+\xi)P^{xx}}\right)_i, \text{ for } i = N+1,\ldots,2N \tag{A.18}$$

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 $(N+\xi)P^{xx}$, $\rho$ is a positive scaling parameter that minimizes higher order effects and can be chosen to be arbitrarily small, $\lambda$ is a positive parameter that guarantees positive-definiteness of the covariance matrix, $\theta$ 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 nonlinear transformation is applied to each of the sigma points defined in $(A.16) - (A.18)$ :

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

The UKF 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 is defined by including the state and measurement noises. This yields 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:

$$\begin{aligned} \mathcal{X}^a_{t|t} &= x_{t|t}, \\ \mathcal{X}^a_{t|t} &= x_{t|t} + \left(\sqrt{(N_a + \xi) P^a_{t|t}}\right)_i, \text{ for } i = 1, \ldots, N_a \\ \mathcal{X}^a_{t|t} &= x_{t|t} - \left(\sqrt{(N_a + \xi) P^a_{t|t}}\right)_i, \text{ for } i = N_a + 1, \ldots, 2N_a \end{aligned} \tag{A.19}$$

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.

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

and

$$P^a_{0|0} = \begin{bmatrix} P_{0|0} & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & R \end{bmatrix}.$$

## A.6 The Particle Filter

This appendix provides a general introduction to particle filters following Johannes and Polson (2009). As mentioned in Chapter 2 the solution to the filtering problem is the conditional distribution of the state variable, $x_t$, given the observed data, $y_1, ..., y_t$. To estimate the density, it is necessary to follow a two-step procedure of prediction and updating. The prediction step combines the current filtering distribution with the state transition,

$$p\left(x_{t+1}\left|y_1, ..., y_t\right.\right) = \int p\left(x_{t+1}\left|x_t\right.\right) p\left(x_t\left|y_1, ..., y_t\right.\right) dx_t, \tag{A.20}$$

providing a forecast of next period's state. Now, given a new observation, $y_{t+1}$, the predictive views are updated by Bayes' rule

$$\begin{aligned} p\left(x_{t+1}\left|y_1, ..., y_{t+1}\right.\right) &= \frac{p\left(y_{t+1}\left|x_{t+1}\right.\right) p\left(x_{t+1}\left|y_1, ..., y_t\right.\right)}{p\left(y_1, ..., y_{t+1}\right)} \\ &\propto p\left(y_{t+1}\left|x_{t+1}\right.\right) p\left(x_{t+1}\left|y_1, ..., y_t\right.\right), \end{aligned} \tag{A.21}$$

where $\propto$ represents that the denominator of Eq.(A.21) is a normalizing constant that does not depend on $x_{t+1}$. The main complication is that the density $p\left(x_t\left|y_1, ..., y_t\right.\right)$ is known analytically only in a few cases, such as the linear model with normally distributed errors. In this case, the standard Kalman filter can be applied to generate the first and second moments of the posterior distribution. In nonlinear and non-normal models, the density $p\left(x_t\left|y_1, ..., y_t\right.\right)$ is a function of the observables, $y_1, ..., y_t$ that has no analytical tractability. Therefore, Monte Carlo methods are a feasible way to estimate $p\left(x_t\left|y_1, ..., y_t\right.\right)$.

### A.6.1 Implementation

A particle filter is a discrete approximation, denoted by $p^N\left(x_t\left|y_1, ..., y_t\right.\right)$, to the conditional density $p\left(x_t\left|y_1, ..., y_t\right.\right)$. The density is generally written as $\left\{\pi_t^{(i)}, x_t^{(i)}\right\}_{i=1}^N$, where the weights add up to one, $\sum_{i=1}^N \pi_t^{(i)} = 1$, and the support of this density is denoted by $x_t^{(i)}$. The generic particle approximation is given by

$$p^N\left(x_t\left|y_1, ..., y_t\right.\right) = \sum_{i=1}^N \pi_t^{(i)} \delta\left(x_t - x_{t-1}^{(i)}\right), \tag{A.22}$$

where $\delta\left(\cdot\right)$ denotes the Dirac delta function.

Intuitively, a particle filter is a discrete approximation to $p\left(x_t\left|y_1, ..., y_t\right.\right)$ that consists of states or *particles*, denoted by $\left\{x_t^{(i)}\right\}_{i=1}^N$, and weights associated to those particles $\left\{\pi_t^{(i)}\right\}_{i=1}^N$. This approximation can be thought as a random histogram, where the states define the support and the weights are the "probabilities"associated with each state.

Let $p^N(x_{t+1}|y_1,...,y_t)$ and $p^N(x_{t+1}|y_1,...,y_{t+1})$ denote the discrete approximation to the conditional densities, $p(x_{t+1}|y_1,...,y_t)$ and $p(x_{t+1}|y_1,...,y_{t+1})$, respectively. Suppose that the conditional densities, $p(x_{t+1}|x_t)$ and $p(y_{t+1}|x_{t+1})$ are known, then by substituting (A.22) in (A.20), and due to the properties of the Dirac delta function, we have that the integral becomes a sum, that is

$$\begin{aligned} p^N(x_{t+1}|y_1,...,y_t) &= \int p(x_{t+1}|x_t)\, p^N(x_t|y_1,...,y_t)\, dx_t \\ &= \sum_{i=1}^{N} p\left(x_{t+1}\left|x_t^{(i)}\right.\right)\pi_t^{(i)}. \end{aligned} \tag{A.23}$$

Finally, substituting $(A.23)$ in Eq.$(A.21)$ yields to

$$p^N(x_{t+1}|y_1,...,y_{t+1}) \propto p(y_{t+1}|x_{t+1}) \sum_{i=1}^{N} p\left(x_{t+1}\left|x_t^{(i)}\right.\right)\pi_t^{(i)}. \tag{A.24}$$

Given the discrete approximation to the conditional density function, $p^N(x_t|y_1,...,y_t)$, the main challenge of particle filtering is to generate a sample from $p^N(x_{t+1}|y_1,...,y_{t+1})$ recursively, after receiving a new observation, $y_{t+1}$.

The particle approximation can be transformed into an equally weighted random sample from the density, $p^N(x_t|y_1,...,y_t)$, by sampling with replacement from the discrete distribution, $\left\{\pi_t^{(i)}, x_t^{(i)}\right\}_{i=1}^{N}$. This procedure is known as resampling and produces a new sample with uniformly distributed weights, i.e. $\pi_t^{(i)} = 1/N$. Resampling can be done in different ways, but the simplest is the multinomial sampling (Casella and Robert, 2004).

### A.6.2 Sampling Importance Resampling

One of the most popular and most general particle filtering algorithm is known as the sampling importance resampling (SIR) algorithm. The algorithm relies on two steps:

**Algorithm A.6.1** *Given samples from* $p(x_t|y_1,...,y_t)$,

1. *Draw* $x_{t+1}^{(i)} \sim p\left(x_{t+1}\left|x_t^{(i)}\right.\right)$ *for* $i = 1,...,N$,
2. *Draw* $z^{(i)} \sim Mult_N\left(\left\{w_{t+1}^{(i)}\right\}_{i=1}^{N}\right)$ *for* $i = 1,...,N$ *and set* $x_{t+1}^{(i)} = x_{t+1}^{z(i)}$

   *where* $Mult_N$ *denotes an* $N-$*component multinomial distribution and the importance weights are given by*

$$w_{t+1}^{(i)} = \frac{p\left(y_{t+1}\left|x_{t+1}^{(i)}\right.\right)}{\sum_{i=1}^{N} p\left(y_{t+1}\left|x_{t+1}^{(i)}\right.\right)}.$$

Prior to resampling, each particle had weight $w_{t+1}^{(i)}$. After resampling, the weights are equal, by the definition of resampling. The intuition of the algorithm is as follows: in the first step, the algorithm simulates new particles from the distribution, $p\left(x_{t+1}\left|x_t^{(i)}\right.\right)$. Upon observing $y_{t+1}$, the resampling step selects the particles that were most likely, in terms of the conditional likelihood, $p\left(y_{t+1}\left|x_{t+1}^{(i)}\right.\right)$, to have generated $y_{t+1}$.

From Chapter 2, we learn that a number of applications characterize the state transition and measurement densities through a state-space model representation of the form:

$$y_t = h\left(x_t\right) + v_t, \qquad v_t \sim \mathcal{N}\left(0, R\right),$$

$$x_{t+1} = g\left(x_t\right) + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim \mathcal{N}\left(0, Q\right).$$

In this case, SIR algorithm becomes:

**Algorithm A.6.2** *Given samples from* $p\left(x_t\left|y_1, ..., y_t\right.\right)$, *denoted by* $x_t^{(i)}$ *for* $i = 1, ..., N$,

1. *Draw* $x_{t+1}^{(i)} \sim\sim \mathcal{N}\left(g\left(x_t^{(i)}\right), Q\right)$ *for* $i = 1, ..., N$,
2. *Draw* $z^{(i)} \sim Mult_N\left(\left\{w_{t+1}^{(i)}\right\}_{i=1}^N\right)$ *for* $i = 1, ..., N$ *and set* $x_{t+1}^{(i)} = x_{t+1}^{z(i)}$

   *where* $Mult_N$ *denotes an* $N-$*component multinomial distribution and the importance weights are given by*

$$w_{t+1}^{(i)} = \frac{\exp\left(-\frac{1}{2}\left(y_{t+1} - h\left(x_{t+1}^{(i)}\right)\right)^\top R^{-1}\left(y_{t+1} - h\left(x_{t+1}^{(i)}\right)\right)\right)}{\sum_{i=1}^N \exp\left(-\frac{1}{2}\left(y_{t+1} - h\left(x_{t+1}^{(i)}\right)\right)^\top R^{-1}\left(y_{t+1} - h\left(x_{t+1}^{(i)}\right)\right)\right)}.$$

The theoretical justification for these algorithms is the weighted bootstrap algorithm or SIR algorithm, which was designed to simulate posterior distributions, of the form $L\left(x\right)p\left(x\right)$, where $L\left(\cdot\right)$ denotes the likelihood function and $p\left(\cdot\right)$ the prior. The algorithm first draws an independent sample $x^{(i)} \sim p\left(x\right)$ for $i = 1, ..., N$, and then computes the normalized importance weights $w^{(i)} = \frac{L\left(x^{(i)}\right)}{\sum_{i=1}^N L\left(x^{(i)}\right)}$. The sample drawn from the discrete distribution $\left\{x^{(i)}, w^{(i)}\right\}_{i=1}^N$ tends in distribution to a sample from the product density $L\left(x\right)p\left(x\right)$ as $N$ increases. [67]

[67] See Casella and Robert (2004) for a detailed explanation.

## A.7 Quasi-Maximum Likelihood Standard Errors

Gallant and White (1988) show that under certain regularity conditions, a heteroskedasticity and autocorrelation consistent covariance matrix of the quasi-maximum likelihood (QML) estimator $\theta^{QML}$ can be obtained using the formula

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

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

$$A_T\left(\theta^{QML}\right) = \frac{\partial^2}{\partial\theta\partial\theta^\top}\mathcal{L}\left(\theta\right),$$

and $B_T$ is a consistent estimator of the covariance matrix of the first derivative of the QML function (2.26). Newey and West (1987) proposed an estimator for $B_T$ given by

$$B_T = \sum_{t=1}^{T} s_t {s_t}^\top + \sum_{t=1}^{L}\sum_{r=t+1}^{T}\left(1 - \frac{t}{L+1}\right)\left[s_t s_{t-r}^\top + s_{t-r}s_t^\top\right],$$

where

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

and $L$ represents the number of sample autocovariances to include in the estimation of the variance–covariance matrix. Newey and West (1994) provide a non-parametric method for automatically selecting $L$ as a function of the number of observations.

## A.8 Calculation of Moments

**Definition** Let $X = (x_1, x_2, x_3, ..., x_N)^\top$ be a normally distributed vector with mean vector $\mu$ and variance covariance matrix $\Sigma$; 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 A.8.1** *Let $X = (x_1, x_2, x_3, ..., x_N)^\top$ 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 A.8.2** *Let $X = (x_1, x_2, x_3)^\top$ be a normally distributed random vector with mean vector $\mu$ and variance covariance matrix, $\Sigma$; 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 Lemma A.8.1.

∎

# Appendix B

# Appendix to Chapter 3

## B.1 Sharpe Ratios in Asset Pricing

Harrison and Kreps (1979) show that the absence of arbitrage implies the existence of a stochastic discount factor (SDF) or pricing kernel, denoted by $M_t$, that prices all assets in the economy.[68] More specifically, the conditional expectation of the product of the stochastic discount factor and the gross asset return ($R_t$) must be equal to one; that is,

$$\mathbb{E}_t \left[ M_{t+1} R_{t+1} \right] = 1, \tag{B.1}$$

where the conditional expectation is based on the information available at time $t$. Since Eq. $(B.1)$ holds for any asset in the economy, it must hold for the one-period risk-free interest rate $(R_{ft+1})$; consequently, the risk-free rate can be written as the inverse of the conditional expectation of the stochastic discount factor,

$$R_{f,t+1} = \frac{1}{\mathbb{E}_t \left[ M_{t+1} \right]}. \tag{B.2}$$

Another implication of Eq. $(B.1)$ is that the expected risk premium on any asset is given by the negative of the product of the risk-free rate and the conditional covariance of the stochastic discount factor with the gross return; that is,

$$\mathbb{E}_t \left[ R_{t+1} - R_{ft+1} \right] = -R_{ft+1} Cov_t \left( R_{t+1}, M_{t+1} \right). \tag{B.3}$$

The conditional Sharpe ratio of an asset at time $t$, denoted by $SR_t$, is defined as the ratio of the conditional mean excess return to the conditional standard deviation of its return; that is,

$$SR_t = \frac{\mathbb{E}_t \left[ R_{t+1} - R_{ft+1} \right]}{\sigma_t \left[ R_{t+1} - R_{ft+1} \right]}. \tag{B.4}$$

[68] See Back (2010) for a detailed and concise explanation.

Then, the conditional Sharpe ratio is proportional to the risk-free rate, the volatility of the pricing kernel and the correlation between the pricing kernel and the return; that is,

$$\frac{\mathbb{E}_t\left[R_{t+1}-R_{ft+1}\right]}{\sigma_t\left[R_{t+1}-R_{ft+1}\right]}=-R_{ft+1}\sigma_t\left[M_{t+1}\right]Corr_t\left[R_{t+1},M_{t+1}\right], \tag{B.5}$$

where $\sigma_t$ and $Corr_t$ are the standard deviation and correlation; respectively, both conditional on information at time $t$. The conditional Sharpe ratio of any asset in the economy is time varying as long as the risk-free rate varies or the pricing kernel is conditionally heteroskedastic; that is, if $\sigma_t\left[M_{t+1}\right]$ changes over time or if the correlation between the stock market return and the stochastic discount factor is time varying.

Now, the maximum of the right-hand side of Eq. $(B.5)$ over all returns defines a lower bound for the standard deviation of any stochastic discount factor depending on the risk-free rate. Since the correlation coefficient is between -1 and 1, we have

$$\frac{\mathbb{E}_t\left[R_{t+1}\right]-R_{ft+1}}{\sigma_t\left[R_{t+1}-R_{ft+1}\right]}\leq R_{ft+1}\sigma_t\left[M_{t+1}\right]\equiv SR_t^{\max},\quad \text{for all assets.} \tag{B.6}$$

Eq. $(B.6)$ implies the Hansen and Jagannathan (1991) bound, which is an upper bound to the absolute value of the conditional Sharpe ratios of any asset in the economy, given a specific discount factor. The maximum Sharpe ratio, $SR_t^{\max}$, is achieved if there exists an asset in the economy which is perfectly negatively correlated with $M_{t+1}$. In general, the Sharpe ratios of all the assets in the economy are bounded by the right-hand side of Eq. $(B.6)$ but when markets are complete there exists an asset that achieves the upper bound, and the inequality becomes an equality.[69] Moreover, a very volatile SDF is necessary to understand high Sharpe ratios. The conditional variance of the SDF can be thought of as the variance of the investor's marginal utility of consumption in the next period.[70] Therefore, from Eq. $(B.5)$ we learn that each model has an implication for the dynamic behavior of the market Sharpe ratio, since each model implies a functional form for the SDF.

The use of log-returns is a common practice in the empirical literature. A standard approximation of the Sharpe ratio based on continuously compounded returns is given by

$$SR_t=\frac{\mathbb{E}_t[r_{t+1}]-r_{f,t+1}+\frac{\sigma_t^2[r_{t+1}]}{2}}{\sigma_t\left[r_{t+1}\right]}, \tag{B.7}$$

where $r_{t+1}$ denotes the continuously compounded return of an asset, $r_{f,t+1}$ denotes the continuously compounded risk-free rate and $\sigma_t\left[r_{t+1}\right]$ denotes the standard deviation of the return of an asset. The numerator in Eq. $(B.7)$ includes the Jensen adjustment for log-returns.[71]

[69] A detailed discussion of this result is shown in Lettau and Uhlig (2002).

[70] Hansen and Jagannathan (1991) provide a comprehensive analysis of this bound, allowing for many risky assets and no risk-free asset, and derive implications of the positivity of the stochastic discount factor.

[71] The difference between Eqs. $(B.6)$ and $(B.7)$ is almost negligible for short return horizons, as reported by

## B.2 The Solution to the Long-Run Risks Model

This section provides solutions for the consumption and dividend claim for the Bansal, Kiku, and Yaron (2012a) endowment process,

$$\begin{aligned} \Delta c_{t+1} &= \mu_c + x_t + \sigma_t \eta_{t+1} \\ x_{t+1} &= \rho x_t + \varphi_e \sigma_t e_{t+1} \\ \sigma_{t+1}^2 &= \overline{\sigma}^2 + v\left(\sigma_t^2 - \overline{\sigma}^2\right) + \sigma_w w_{t+1} \\ \Delta d_{t+1} &= \mu_d + \phi x_t + \varphi \sigma_t u_{t+1} + \pi \sigma_t \eta_{t+1} \\ w_{t+1}, e_{t+1}, u_{t+1}, \eta_{t+1} &\sim i.i.d.\ \mathcal{N}(0,1). \end{aligned} \tag{B.8}$$

The Euler equation for this economy is

$$\mathbb{E}_t\left[\exp\left(\theta \ln \delta - \frac{\theta}{\psi}\Delta c_{t+1} + (\theta - 1) r_{a,t+1} + r_{i,t+1}\right)\right] = 1, \tag{B.9}$$

where $r_{a,t+1}$ is the log-return on the consumption claim and $r_{i,t+1}$ is the log-return on any asset. All returns are given by the approximation from Campbell and Shiller (1988), $r_{i,t+1} = \kappa_{0,i} + \kappa_{1,i} z_{i,t+1} - z_{i,t} + \Delta d_{i,t+1}$.

Let $Y_t^\top = [1, x_t, \sigma_t^2]$ denote a vector of state variables and the log price-consumption ratio be given by $z_t = A^\top Y_t$, where $A$ denotes a vector of coefficients $A^\top = [A_0, A_1, A_2]$. In general, for any other asset $i$, define the coefficients in the same manner: $A_i^\top = [A_{0,i}, A_{1,i}, A_{2,i}]$. This section calculates the price of the consumption claim as well as the dividend claim $z_{t,m} = A_m^\top Y_t$. The coefficients that characterize $z_t$ and $z_{t,m}$ are obtained by the method of undetermined coefficients and by the fact that the Euler equation must hold for all values of $Y_t^\top$.

The risk premium on any asset is

$$\begin{aligned} \mathbb{E}_t\left[r_{i,t+1} - r_{f,t}\right] + \frac{1}{2} Var_t\left[r_{i,t+1}\right] &= -Cov_t\left(m_{t+1}, r_{i,t+1}\right) \\ &= \sum_{j=n,e,w} \lambda_j \beta_{i,j} \sigma_{j,t}^2, \end{aligned} \tag{B.10}$$

where $\beta_{i,j}$ is the beta and $\sigma_{j,t}^2$ the volatility of the $j^{th}$ risk source, and the $\lambda_j$ represents the price of each risk source.

Brandt and Kang (2004). Nielsen and Vassalou (2004) analyze the difference between discrete and continuously compounded versions of Sharpe ratios and propose this adjustment for performance evaluation. Campbell and Viceira (2002) discuss in detail this approximation in a portfolio optimization framework.

### B.2.1 Consumption Claim

The risk premium for the consumption claim is

$$\mathbb{E}_t\left[r_{a,t+1} - r_{f,t}\right] + \frac{1}{2}Var_t\left[r_{a,t+1}\right] = \lambda_n \beta_{a,n}\sigma_t^2 + \lambda_e \beta_{a,e}\sigma_t^2 + \lambda_w \beta_{a,w}\sigma_w^2, \tag{B.11}$$

where $\beta_{a,n} = 1$, $\beta_{a,e} = \kappa_1 A_1 \varphi_e$ and $\beta_{a,w} = \kappa_1 A_2$. The conditional variance of the consumption claim is equal to

$$Var_t\left[r_{a,t+1}\right] = \left(\beta_{a,n}^2 + \beta_{a,e}^2\right)\sigma_t^2 + \beta_{a,w}^2\sigma_w^2. \tag{B.12}$$

The coefficients $A'$ for the log price-consumption ratio $z_t$ are

$$\begin{aligned} A_0 &= \frac{\ln\delta + \mu_c\left(1 - \frac{1}{\psi}\right) + \kappa_0 + \beta_{a,w}\overline{\sigma}^2\left(1 - v\right) + \frac{1}{2}\theta\beta_{a,w}^2\sigma_w^2}{\left(1 - \kappa_1\right)}, \\ A_1 &= \frac{1 - \frac{1}{\psi}}{1 - \kappa_1\rho}, \\ A_2 &= \frac{\frac{\theta}{2}\left[\left(1 - \frac{1}{\psi}\right)^2 + \beta_{a,e}^2\right]}{\left(1 - \kappa_1 v_1\right)}. \end{aligned} \tag{B.13}$$

### B.2.2 Dividend Claim

The innovation to the market return, denoted by $r_{m,t+1} - \mathbb{E}_t\left(r_{m,t+1}\right)$, is

$$r_{m,t+1} - \mathbb{E}_t\left(r_{m,t+1}\right) = \varphi\sigma_t u_{t+1} + \beta_{m,\eta}\sigma_t\eta_{t+1} + \beta_{m,e}\sigma_t e_{t+1} + \beta_{m,w}\sigma_w w_{t+1}, \tag{B.14}$$

where $\beta_{m,\eta} = \pi$, $\beta_{m,e} = \kappa_{1,m}A_{1,m}\varphi_e$ and $\beta_{m,w} = \kappa_{1,m}A_{2,m}$, which implies that the risk premium on the dividend claim is

$$\mathbb{E}_t\left[r_{m,t+1} - r_{f,t}\right] + \frac{1}{2}Var_t\left[r_{m,t+1}\right] = \lambda_\eta \beta_{m,\eta}\sigma_t^2 + \lambda_e \beta_{m,e}\sigma_t^2 + \lambda_w \beta_{m,w}\sigma_w^2. \tag{B.15}$$

Finally, the coefficients $A'_m$ for the log price-dividend ratio are as follows

$$\begin{aligned} A_{0,m} &= \frac{\begin{bmatrix} \theta\ln\delta + \mu_c\left(\theta - \frac{\theta}{\psi} - 1\right) - \lambda_w\overline{\sigma}^2\left(1 - v\right) + \left(\theta - 1\right)\left[\kappa_0 + A_0\left(\kappa_1 - 1\right)\right] \\ \kappa_{0,m} + \beta_{m,w}\overline{\sigma}^2\left(1 - v\right) + \mu_d + \frac{1}{2}\left[\beta_{m,w} - \lambda_w\right]^2\sigma_w^2 \end{bmatrix}}{\left(1 - \kappa_{1,m}\right)}, \\ A_{1,m} &= \frac{\phi - \frac{1}{\psi}}{1 - \kappa_{1,m}\rho}, \\ A_{2,m} &= \frac{\left(1 - \theta\right)A_2\left(1 - \kappa_1 v_1\right) + \frac{1}{2}\left[\left(\pi - \lambda_n\right)^2 + \left[\beta_{m,e} - \lambda_e\right]^2 + \varphi^2\right]}{\left(1 - \kappa_{1,m}v\right)}. \end{aligned} \tag{B.16}$$

### B.2.3 Risk-Free Interest Rate

The risk-free rate is derived from the Euler equation applied to the risk-less asset:

$$\begin{aligned} r_{f,t+1} &= -\log \mathbb{E}_t \left[\exp\left(m_{t+1}\right)\right] \\ &= -\theta \ln \delta + \frac{\theta}{\psi}\mathbb{E}_t\left[\Delta c_{t+1}\right] + (1-\theta)\,\mathbb{E}_t\left[r_{a,t+1}\right] \\ &\quad -\frac{1}{2}Var_t\left[\frac{\theta}{\psi}\Delta c_{t+1} + (1-\theta)\, r_{a,t+1}\right]. \end{aligned} \tag{B.17}$$

By subtracting $(1-\theta)\, r_{f,t+1}$ from both sides of Eq. $(B.17)$ and if $\theta \neq 0$, then we can divide by $\theta$, yielding to an expression for the risk-free rate

$$r_{f,t+1} = -\ln \delta + \frac{1}{\psi}\mathbb{E}_t\left[\Delta c_{t+1}\right] + \frac{(1-\theta)}{\theta}\mathbb{E}_t\left[r_{a,t+1} - r_{f,t+1}\right] - \frac{1}{2\theta}Var_t\left(m_{t+1}\right), \tag{B.18}$$

where

$$\begin{aligned} \mathbb{E}_t\left[\Delta c_{t+1}\right] &= \mu_c + x_t \\ \mathbb{E}_t\left[r_{a,t+1} - r_{f,t+1}\right] &= \left(\lambda_n \beta_{a,n} + \lambda_e \beta_{a,e} - \frac{\left(\beta_{a,n}^2 + \beta_{a,e}^2\right)}{2}\right)\sigma_t^2 + \left(\lambda_w \beta_{a,w} - \frac{\beta_{a,w}^2}{2}\right)\sigma_w^2. \\ Var_t\left(m_{t+1}\right) &= \left(\lambda_\eta^2 + \lambda_e^2\right)\sigma_t^2 + \lambda_w^2 \sigma_w^2, \end{aligned}$$

as a result

$$r_{f,t+1} = A_{0,f} + A_{1,f} x_t + A_{2,f}\sigma_t^2, \tag{B.19}$$

where

$$\begin{aligned} A_{0,f} &= -\ln \delta + \frac{\mu_c}{\psi} + \frac{(1-\theta)}{\theta}\left(\lambda_w \beta_{a,w} - \frac{\beta_{a,w}^2}{2}\right)\sigma_w^2 - \frac{\lambda_w^2 \sigma_w^2}{2\theta}, \\ A_{1,f} &= \frac{1}{\psi}, \\ A_{2,f} &= \frac{(1-\theta)}{\theta}\left(\lambda_n \beta_{a,n} + \lambda_e \beta_{a,e} - \frac{\left(\beta_{a,n}^2 + \beta_{a,e}^2\right)}{2}\right) - \frac{\left(\lambda_\eta^2 + \lambda_e^2\right)}{2\theta}. \end{aligned}$$

### B.2.4 Return on the Market Portfolio

Recall that the rate of return on the market portfolio is

$$r_{m,t+1} = \kappa_{0,m} + \kappa_{1,m} z_{m,t+1} - z_{m,t} + \Delta d_{t+1}, \tag{B.20}$$

where the dynamics are characterized by the following equations

$$\begin{aligned} z_{m,t} &= A_{0,m} + A_{1,m}x_t + A_{2,m}\sigma_t^2 \\ x_{t+1} &= \rho x_t + \varphi_e \sigma_t e_{t+1} \\ \sigma_{t+1}^2 &= \overline{\sigma}^2 + v\left(\sigma_t^2 - \overline{\sigma}^2\right) + \sigma_w w_{t+1}, \\ \Delta d_{t+1} &= \mu_d + \phi x_t + \varphi \sigma_t u_{t+1} + \pi \sigma_t \eta_{t+1}. \end{aligned} \tag{B.21}$$

Now, since each of the components of the market return follows a normal distribution, then the market return has a normal distribution with conditional mean

$$\begin{aligned} \mathbb{E}_t\left[r_{m,t+1}\right] &= \kappa_{0,m} + \kappa_{1,m}\mathbb{E}_t\left[z_{m,t+1}\right] - z_{m,t} + \mathbb{E}_t\left[\Delta d_{t+1}\right] \\ &= \kappa_{0,m} + \kappa_{1,m}\left(A_{0,m} + A_{1,m}\rho x_t + A_{2,m}\left(\overline{\sigma}^2 + v\left(\sigma_t^2 - \overline{\sigma}^2\right)\right)\right) \\ &\quad -A_{0,m} - A_{1,m}x_t - A_{2,m}\sigma_t^2 + \mu_d + \phi x_t \\ &= \kappa_{0,m} + \left(\kappa_{1,m} - 1\right)A_{0,m} + \kappa_{1,m}A_{2,m}\left(1 - v\right)\overline{\sigma}^2 + \mu_d \\ &\quad + \left[A_{1,m}\left(\kappa_{1,m}\rho - 1\right) + \phi\right]x_t + A_{2,m}\left(\kappa_{1,m}v - 1\right)\sigma_t^2 \\ &= B_0 + B_1 x_t + B_2 \sigma_t^2, \end{aligned} \tag{B.22}$$

where

$$\begin{aligned} B_0 &= \kappa_{0,m} + \left(\kappa_{1,m} - 1\right)A_{0,m} + \kappa_{1,m}A_{2,m}\left(1 - v\right)\overline{\sigma}^2 + \mu_d \\ B_1 &= \phi - A_{1,m}\left(1 - \kappa_{1,m}\rho\right) = \frac{1}{\psi}, \\ B_2 &= A_{2,m}\left(\kappa_{1,m}v - 1\right). \end{aligned}$$

Now, the variance of the market portfolio is given by

$$\begin{aligned} Var_t\left[r_{m,t+1}\right] &= \kappa_{1,m}^2 Var_t\left[z_{m,t+1}\right] + Var_t\left[\Delta d_{t+1}\right] \\ &= \kappa_{1,m}^2\left(A_{1,m}^2\varphi_e^2\sigma_t^2 + A_{2,m}^2\sigma_w^2\right) + \left(\varphi^2 + \pi^2\right)\sigma_t^2 \\ &= D_0 + D_1\sigma_t^2, \end{aligned}$$

where $D_0 = \left(\kappa_{1,m}A_{2,m}\sigma_w\right)^2$ and $D_1 = \kappa_{1,m}^2 A_{1,m}^2\varphi_e^2 + \varphi^2 + \pi^2$.

### B.2.5 Linearization Parameters

For any asset, the linearization parameters are determined endogenously by the following system of equations as discussed in Bansal, Kiku, and Yaron (2012a) and Beeler and Campbell (2012):

$$\begin{aligned} \overline{z_i} &= A_{0,i}\left(\overline{z_i}\right) + A_{2,i}\left(\overline{z_i}\right)\sigma^2, \\ \kappa_{1,i} &= \frac{\exp\left(\overline{z_i}\right)}{1+\exp\left(\overline{z_i}\right)}, \\ \kappa_{0,i} &= \ln\left(1+\exp\left(\overline{z_i}\right)\right) - \kappa_{1,i}\overline{z_i}. \end{aligned} \tag{B.23}$$

The solution is determined numerically by iteration until reaching a fixed point of $\overline{z_i}$. The dependence of $A_{0,i}$ and $A_{2,i}$ on the linearization parameters has been discussed in previous sections.

## B.3 Excess Returns Conditional Moments Implied by the Long-Run Risks Model

### B.3.1 Expected Returns

The expected excess returns for period $k$ are defined as

$$r_{m,t+k+1} - r_{f,t+k+1},\ k = 0, 1, 2, ...$$

Now, the conditional excess risk premium for any period has a closed-form expression given by

$$\mathbb{E}_t \left[ r_{m,t+k+1} - r_{f,t+k+1} \right] = E_{0,k+1} + E_{1,k+1}\sigma_t^2, \tag{B.24}$$

where

$$\begin{aligned} E_{0,k+1} &= E_0 + E_1 \left(1 - v^k\right) \overline{\sigma}^2, \\ E_{1,k+1} &= E_1 v^k,\ k = 0, 1, 2, ..., \\ E_0 &= B_0 - A_{0,f}, \\ E_1 &= B_2 - A_{2,f}. \end{aligned}$$

### B.3.2 Variance of Excess Returns

Now, for any time period $k$, the conditional variance of the future excess returns is given by

$$Var_t \left[ r_{m,t+k+1} - r_{f,t+k+1} \right],\ \text{for } k = 0, 1, 2, ...$$

Its closed-form expression is given by

$$Var_t \left[ r_{m,t+k+1} - r_{f,t+k+1} \right] = D_{0,k+1} + D_{1,k+1}\sigma_t^2, \tag{B.25}$$

where

$$\begin{aligned} D_{0,k+1} &= D_0 + D_1 \left(1 - v^k\right) \overline{\sigma}^2 + E_1^2 \sigma_w^2 \frac{1 - v^{2k}}{1 - v^2} \\ D_{1,k+1} &= v^k D_1, \\ D_0 &= \left(\kappa_{1,m} A_{2,m} \sigma_m\right)^2, \\ D_1 &= \left(\kappa_{1,m} A_{1,m} \varphi_e\right)^2 + \varphi^2 + \pi^2. \end{aligned}$$

**Autocovariance of Excess Returns**

Now, let $0 \leq k < p$. Then the autocovariance of excess returns is

$$cov_t\left(r_{m,t+k+1} - r_{f,t+k+1}, r_{m,t+p+1} - r_{f,t+p+1}\right) = E_1^2 \sigma_w^2 v^{p-k} \left(\frac{1-v^{2k}}{1-v^2}\right) + E_1 \kappa_{1,m} A_{2,m} \sigma_w^2 v^{p-k-1}.$$

### B.3.3 Aggregate Excess Returns

Now, the expected excess returns during $K$ periods are given by the sum of the one-period excess returns,

$$\sum_{k=1}^{K} \left(r_{m,t+k} - r_{f,t+k}\right).$$

Its conditional mean is

$$\mathbb{E}_t \left[\sum_{k=1}^{K} r_{m,t+k} - r_{f,t+k}\right] = \mathbf{E}_{0,K} + \mathbf{E}_{1,K} \sigma_t^2,$$

where

$$\begin{aligned} \mathbf{E}_{0,K} &= K E_0 + E_1 \overline{\sigma}^2 \left[K - \frac{\left(1-v^K\right)}{(1-v)}\right], \\ \mathbf{E}_{1,K} &= E_1 \frac{\left(1-v^K\right)}{(1-v)}. \end{aligned}$$

### B.3.4 Variance of Aggregate Excess Returns

The conditional variance is

$$Var_t \left[\sum_{k=1}^{K} r_{m,t+k} - r_{f,t+k}\right] = \mathbf{D}_{0,K} + \mathbf{D}_{1,K} \sigma_t^2.$$

where

$$\begin{aligned}
\mathbf{D}_{0,K} &= KD_0 + KD_1\overline{\sigma}^2 - D_1\overline{\sigma}^2\frac{\left(1-v^K\right)}{(1-v)} \\
&\quad + \frac{E_1^2\sigma_w^2}{1-v^2}\left[K - \frac{\left(1-v^{2K}\right)}{(1-v^2)}\right] \\
&\quad + 2\left[K - \frac{1-v^K}{1-v}\right]\left[\begin{array}{c} \frac{E_1^2\sigma_w^2}{(1-v^2)}\left[\frac{v}{1-v}\right] \\ +E_1\kappa_{1,m}A_{2,m}\sigma_w^2\left[\frac{1}{1-v}\right] \end{array}\right] \\
&\quad - \frac{2E_1^2\sigma_w^2}{(1-v^2)}\frac{v^3\left(1-v^{K-1}\right)\left(1-v^K\right)}{(1-v)^2(1+v)}, \\
\mathbf{D}_{1,K} &= \frac{\left(1-v^K\right)}{(1-v)}D_1.
\end{aligned}$$

## B.4 Quasi-Maximum Likelihood Estimation

Since the measurement equation considered in each of the models is nonlinear, one possibility is to rely on Taylor series approximations to obtain extended forms of the Kalman filter. The transition and measurement equations analyzed in the previous section are expressed as follows:

$$y_t = \mu(x_{t-1}) + \lambda(x_{t-1})\,\varepsilon_t, \tag{B.26}$$

$$x_t = Ax_{t-1} + \eta_t, \tag{B.27}$$

where $\varepsilon_t$ follows a standard normal distribution and $\eta_t$ is a $d$-dimensional noise vector with variance–covariance matrix $\Sigma$. The deterministic functions $\mu(x_t)$ and $\lambda(x_t)$ define the conditional mean and volatility of excess returns and are characterized by each of the models.

I use Gaussian approximations to filter the mean and covariance of the states and measurement series. More specifically, the linearity of the state vector implies that the first and second conditional moments of the state vectors are

$$\begin{aligned} x_{t+1|t} &= Ax_{t|t}\,, & \text{(B.28)}\\ P_{t+1|t} &= AP_{t|t}A^{\top} + \Sigma, & \text{(B.29)} \end{aligned}$$

where $x_{t+1|t}$ and $P_{t+1|t}$ are the time $t$ predicted values of the conditional mean and covariance matrix of the state vector, respectively. These moments allow us to generate a predicted mean $y_{t+1|t}$ and covariance matrix $P^{yy}_{t+1|t}$ of the measurement series, given by

$$\begin{aligned} y_{t+1|t} &= \mathbb{E}\left[\mu(x_t) + \lambda(x_t)\,\varepsilon_{t+1}\,|y_t, y_{t-1}..., y_0\right], & \text{(B.30)}\\ P^{yy}_{t+1|t} &= Var\left[\mu(x_t) + \lambda(x_t)\,\varepsilon_{t+1}\,|y_t, y_{t-1}, ..., y_0\right]. \end{aligned}$$

Finally, the covariance between the observed and unobserved variables, $P^{xy}_{t+1|t}$, is

$$P^{xy}_{t+1|t} = Cov\left[x_{t+1}, \mu(x_t) + \lambda(x_t)\,\varepsilon_{t+1}\,|y_t, y_{t-1}, ..., y_0\right]. \tag{B.31}$$

Using these conditional moments, we apply the Kalman update, represented by the following set of recursive equations to obtained values for the conditional mean $S_{t+1|t+1}$ and covariance $P_{t+1|t+1}$:

$$\begin{aligned} K_{t+1} &= P^{xy}_{t+1|t}\left(P^{yy}_{t+1|t}\right)^{-1}, & \text{(B.32)}\\ x_{t+1|t+1} &= x_{t+1|t} + K_{t+1}\left(y_{t+1} - y_{t+1|t}\right), \\ P_{t+1|t+1} &= P_{t+1|t} - K_{t+1}P^{yy}_{t+1|t}K^{\top}_{t+1}. \end{aligned}$$

The first attempt to estimate the moments in Eqs. $(B.30)$ through $(B.32)$ uses closed-form expression, if available. An alternative way to solve the problem is to use Taylor series expansions of $\mu(x_t)$ and $\lambda(x_t)$ around $x_{t+1|t}$, for an arbitrary number of terms. Properties of this method as well as a detailed explanation can be found in Chapter 2 of this work.

## B.5 External Habit Formation Model

This section presents the model by Campbell and Cochrane (1999) in discrete time and its extension in Wachter (2005). A representative investor is assumed to have state-dependent preferences. More specifically, an investor has utility over consumption relative to a reference point $X_t$ and maximizes

$$\mathbb{E}\left[\sum_{t=0}^{\infty}\delta^t\frac{(C_t-X_t)^{1-\gamma}-1}{1-\gamma}\right], \tag{B.33}$$

where $\delta>0$ is the time preference parameter and $\gamma>0$ is the curvature parameter.

Each investor is concerned with her consumption relative to that of others. Habit $X_t$ is defined through surplus consumption $S_t$, where

$$S_t\equiv\frac{C_t-X_t}{C_t}. \tag{B.34}$$

One can interpret $S_t$ as a business cycle indicator. In economic booms, consumption substantially exceeds the external habit and the surplus, $S_t$, is large; and in recessions consumption barely exceeds the external habit, and the external habit is relatively small.

It is assumed that $s_t=\log S_t$ follows the process

$$s_{t+1}=(1-\phi)\,\overline{s}+\phi s_t+\lambda\left(s_t\right)\left(\Delta c_{t+1}-E_t\left[\Delta c_{t+1}\right]\right), \tag{B.35}$$

where $\overline{s}$ is the unconditional mean of $s_t$, $\phi$ is the persistence and $\lambda\left(s_t\right)$ is the sensitivity of the changes in consumption. The unconditional mean and the sensitivity function are defined in terms of primitive parameters. It is assumed that aggregate consumption growth is log-normal with independent and identically distributed innovations; that is,

$$\Delta c_{t+1}=g+v_{t+1}, \tag{B.36}$$

where $c_t=\log C_t$ and $v_{t+1}\sim N\left(0,\sigma_v^2\right)$ is an i.i.d. sequence. The process for $s_t$ is heteroscedastic and perfectly conditionally correlated with innovations in consumption growth. The sensitivity function $\lambda\left(s_t\right)$ is specified so that the real risk-free rate is linear, and for $s_t\approx\overline{s}$, $x_t$ is a deterministic function of past consumption. Consequently, we have

$$\lambda\left(s_t\right)=\begin{cases}1/\overline{S}\sqrt{1-2\left(s_t-\overline{s}\right)}-1, & \text{if } s_t\leq s_{\max}\\ 0 \quad \text{otherwise,}\end{cases} \tag{B.37}$$

$$\overline{S}=\sigma_v\sqrt{\frac{\gamma}{1-\phi-b/\gamma}}, \tag{B.38}$$

where $b$ is a preference parameter that determines the behavior of the risk-free rate and $s_{\max} = \overline{s} + \frac{1}{2}\left(1 - \overline{S}^2\right)$. In Campbell and Cochrane (1999), $b$ is chosen to be zero and produce a constant real risk-free rate, while Wachter (2005) shows that values of $b > 0$, imply a risk-free rate that is linear in $s_t$.

### B.5.1 Stochastic Discount Factor

Since the habit is external, the investor's inter-temporal marginal rate of substitution is given by

$$M_{t+1} = \delta \left(\frac{S_{t+1}}{S_t}\right)^{-\gamma} \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}. \tag{B.39}$$

Moreover, any asset return $R_{t+1}$ must satisfy

$$\mathbb{E}_t \left[M_{t+1} R_{t+1}\right] = 1. \tag{B.40}$$

### B.5.2 Risk-Free Rate and Maximum Sharpe Ratio

Let $R_{f,t+1}$ denote the one-period risk-free return between $t$ and $t+1$, and $r_{f,t+1} = \ln\left(R_{f,t+1}\right)$; as a result, from Eqs. $(B.39)$ and $(B.40)$ imply that

$$\begin{aligned} r_{f,t+1} &= -\ln\left(\mathbb{E}_t\left[M_{t+1}\right]\right) \\ &= -\ln\left(\delta\right) + \gamma g + \gamma\left(1-\phi\right)\left(\overline{s} - s_t\right) - \frac{\gamma^2 \sigma_v^2}{2}\left(1 + \lambda\left(s_t\right)\right)^2 \\ &= -\ln\left(\delta\right) + \gamma g - \frac{\gamma\left(1-\phi\right) - b}{2} + b\left(\overline{s} - s_t\right), \end{aligned} \tag{B.41}$$

where the last equality comes from substituting the definition of $\lambda\left(s_t\right)$. This definition implies a risk-free rate linear in $s_t$.

Conditional on the information at time $t$, the one-period stochastic discount factor, defined in Eq. $(B.39)$ is the exponential of a normally distributed random variable that has variance $\gamma^2\left[1 + \lambda\left(S_t\right)\right]^2 \sigma^2$. As a result, the Hansen-Jagannathan bound implies that

$$\sqrt{\exp\left(\gamma^2\left[1 + \lambda\left(S_t\right)\right]^2 \sigma^2\right) - 1}$$

is an upper bound on the Sharpe ratio of any portfolio. If $\lambda$ is a decreasing function of $S_t$, then the upper bound on Sharpe ratios will be counter-cyclical: higher in recessions than in booms.

### B.5.3 Price-Dividend Ratio

The aggregate market is represented as the claim to the future consumption stream. If $P_t$ denotes the ex-dividend price of this claim, then Eq. $(B.40)$ implies that in equilibrium $P_t$

satisfies

$$\mathbb{E}_t\left[M_{t+1}\left(\frac{P_{t+1}+C_{t+1}}{P_t}\right)\right]=1, \tag{B.42}$$

which can be rewritten as

$$\mathbb{E}_t\left[M_{t+1}\left(1+\frac{P_{t+1}}{C_{t+1}}\right)\frac{C_{t+1}}{C_t}\right]=\frac{P_t}{C_t}.$$

Since $C_t$ is the dividend paid by the aggregate market, $P_t/C_t$ is the price-dividend ratio. The price-dividend ratio can be computed numerically using numerical methods; Wachter (2005) provides an efficient method for its computation.

Returns on the aggregate market are defined as

$$R^m_{t+1}=\left(\frac{P_{t+1}/C_{t+1}+1}{P_t/C_t}\right)\frac{C_{t+1}}{C_t}.$$

The main difficulty lies in solving the model ($B.42$) for the price-dividend ratio as a function of $s_t$. Once the price-dividend ratio is calculated numerically, Monte Carlo simulations can be performed to obtain accurate estimates of expected returns, volatilities and Sharpe ratios for different holding periods. Details about the simulations are explained in Wachter (2005).