ESSAYS ON ESTIMATION OF NON-LINEAR STATE-SPACE MODELS.

by

Dharmarajan Hariharan

B.E. (Chemical Engineering) , Bangalore University, 1997

M.S. (Chemical Engineering), West Virginia University, 2001

Submitted to the Graduate Faculty of

the Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF ECONOMICS

This dissertation was presented

by

Dharmarajan Hariharan

It was defended on

July 11th 2008

and approved by

Jean-François Richard, Department of Economics

David N. DeJong, Department of Economics

James Feigenbaum, Department of Economics

Roman Liesenfeld, Universität Kiel

Dissertation Advisors: Jean-François Richard, Department of Economics,

David N. DeJong, Department of Economics

ESSAYS ON ESTIMATION OF NON-LINEAR STATE-SPACE MODELS.

Dharmarajan Hariharan, PhD

University of Pittsburgh, 2008

The first chapter of my thesis (co-authored with David N. DeJong, Jean-François Richard and Roman Liesenfeld) develops a numerical procedure that facilitates efficient likelihood evaluation and filtering in applications involving non-linear and non-Gaussian state-space models. These tasks require the calculation of integrals over unobservable state variables. We introduce an efficient procedure for calculating such integrals: the EIS-Filter. The procedure approximates necessary integrals using continuous approximations of target densities. Construction is achieved via efficient importance sampling, and approximating densities are adapted to fully incorporate current information. Extensive comparisons to the standard particle filter are presented using four diverse examples.

The second chapter illustrates the use of copulas to create low-dimensional multivariate importance sampling densities. Copulas enable the problem of multivariate density approximation to be split into a sequence of simpler univariate density approximation problems for the marginals, with the dependence accounted by the copula parameter(s). This separation of the marginals from their dependence allows maximum flexibility in the selection of marginal densities. Combined with the EIS method for refining importance sampling densities, copula densities offer substantial flexibility in creating multivariate importance samplers. In a simulation exercise, we compare the accuracy of the copula-based EIS-Filter to the particle filter in evaluating the likelihood function and in obtaining filtered estimates of the latent variables.

Reliability of growth forecasts critically depend on being able to anticipate/recognize shifts of the economy from recessions to expansions or vice versa. It is widely accepted

that the processes that govern these shifts could be highly non-linear. In the third chapter (co-authored with David N. DeJong, Jean-François Richard and Roman Liesenfeld), we study regime shifts using a non-linear model of GDP growth. The model characterizes growth as following non-linear trajectories that fluctuate stochastically between alternative periods of general acceleration and deceleration. Also, we introduce a non-stochastic rule-based recession-dating method to forecast likely dates for the start of a recession and it length. Results indicate that the model is capable of exhibiting substantially non-linear behavior in its regime-specific latent process and hence is able to anticipate and detect regime-shifts accurately, improving the quality of growth forecasts obtained from it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I am deeply grateful to my advisors Jean-Franois Richard and David N. DeJong for their guidance and encouragement throughout my graduate career. I am also thankful for all the advice and help I have received from Roman Liesenfeld. Special thanks also goes to Martin Burda for helpful comments and discussions. Finally, I would like to thank members of my family, both my parents and my sisters for all the support and encouragement.

1.0 INTRODUCTION

Modern dynamic econometric models often incorporate key latent (unobservable) variables. A wide range of such models can be characterized by an underlying state-space transition model describing the dynamic stochastic process driving the latent variables, in combination with a set of (measurement) equations that link the latent variables to observables. For example, models of real business cycle activity highlight movements of observed output growth in response to innovations in unobserved total factor productivity.

Likelihood-based inference, whether on model parameters or estimates of the latent variables (filtered values), requires integration with respect to the unobservables. Under restrictive assumptions (combining linear equations with Gaussian errors), such integrals can be evaluated analytically using the Kalman filter. Under more flexible assumptions, required integrals must be calculated numerically.

Advances in computing power, coupled with the development of numerical Monte Carlo integration techniques, provide powerful tools for the analysis of high-dimensional non-linear and/or non-Gaussian state-space models. However, when applied to complex dynamic processes, Monte Carlo methods can be highly inefficient because they may require delicate fine tuning as well as very large number of draws to produce numerically accurate estimates of the quantities of interest.

The object of my dissertation is to develop generic, fully automated, robust and numerically efficient numerical Monte Carlo integration techniques that can be successfully applied to high-dimensional non-linear and/or non-gaussian state-space models.

Particle filters represent a landmark development in our ability to evaluate likelihood func-

tions in non-linear and/or non-gaussian state-space models. These methods employ discrete approximations to the densities of the state variables, with the recursive updating of these distributions being a straightforward application of Bayes theorem. The generic simplicity of particle filters explains their growing popularity across a wide range of disciplines. Unfortunately, outliers and other potential pathologies can result in major inefficiencies that require prohibitively large numbers of MC draws to obtain numerically accurate estimates of the likelihood function, and filtered values of the state variables.

In the first chapter we create continuous approximations to the densities of interest using recently developed efficient MC techniques (Efficient Importance Sampling - EIS) to produce substantially superior numerical approximations for relevant integrals. Our method typically produces dramatic reductions in the number of draws required for accurate integration, and is also robust to outliers and other pathological scenarios.

The relative accuracy of our EIS-based filter is demonstrated through four diverse examples. The first involves likelihood evaluation for a model featuring fat-tailed measurement errors. The second involves obtaining filtered estimates of the latent volatility in a stochastic volatility model. The third and the fourth examples highlight pathological situations involving singular distributions and/or the partial observability of data: likelihood inference in an economic growth model with a partly non-stochastic state-process; and filtering in the bearings-only tracking problem.

Since likelihood evaluation in state-space models typically requires high-dimensional integration, a well-established method for overcoming the curse of dimensionality is to decompose the total likelihood into a product of lower-dimensional conditional likelihoods. These conditional likelihoods are then computed sequentially.

The second essay proposes the use of copulas to create low-dimensional multivariate importance sampling densities that allow for maximum flexibility in the selection of marginal densities. Copulas enable the problem of multivariate density approximation to be split into a sequence of simpler univariate density approximation problems for the marginals, with the dependence accounted for by the copula parameter(s). Combined with the EIS method for

refining importance sampling densities, copula densities offer enormous flexibility in creating multivariate importance samplers. For example, while a bivariate Gaussian density requires the marginal densities to be Gaussian themselves, a bivariate meta-Gaussian copula can combine arbitrary marginals via the normal quantile transform to create very general bivariate densities.

The combined strengths of EIS and copula densities are illustrated via an application to a dynamic stochastic general equilibrium (DSGE) model. In a simulation study, we show that the copula-based EIS-Filter is capable of creating accurate importance sampling densities and hence provide accurate likelihood values and filtered estimates of the unobserved variables.

Reliability of growth forecasts critically depend on anticipating/recognizing shifts of the economy from recessions to expansions or vice versa. It is widely accepted that the processes that govern these shifts could be highly non-linear. When combined with the decrease in volatility of observed growth since the mid 1980's (see Stock and Watson, 2002, for a thorough investigation of a broad range of aggregate time series and a comprehensive literature review) the task of predicting regime shifts has only become more challenging.

In the final essay, we study regime shifts using a non-linear model of GDP growth that builds on the work of DeJong, Liesenfeld, and Richard (2005). The model characterizes growth as following non-linear trajectories that fluctuate stochastically between alternative periods of general acceleration and deceleration. Regime changes occur stochastically, with probabilities determined by an observed indicator variable via a logistic link function. We refer to the observed indicator variable as a "tension index". In order to account for the fact that no two regimes are alike, we model the parameters that dictate growth trajectories in each regime as random variables that vary from regime-to-regime. The latent parameters also include the volatility of growth in each regime. This treatment of the volatility parameter as being a latent variable whose distribution is estimated can account for both high and low volatility episodes as naturally arising due to particular realizations of the latent variable.

The increased non-linearity of the latent drift process governing growth in any given regime allows for a better fit of the model and substantially enhanced sensitivity to fluctu-

ations in growth that might be indicative of regime-changes. The distributional characterization of the parameters governing the latent drift process is sufficiently broad to withstand mild trends in the latent parameters. Results from regime-break prediction and in-sample forecasting exercises amply illustrates the benefits of the new model specification over the DeJong, Liesenfeld, and Richard (2005) model.

2.0 AN EFFICIENT APPROACH TO ANALYZING STATE-SPACE REPRESENTATIONS

2.1 INTRODUCTION

Likelihood evaluation and filtering in applications involving state-space models requires the calculation of integrals over unobservable state variables. When models are linear and stochastic processes are Gaussian, required integrals can be calculated analytically via the Kalman filter. Departures entail integrals that must be approximated numerically. Here we introduce an efficient procedure for calculating such integrals: the EIS filter.

The procedure takes as a building block the pioneering approach to likelihood evaluation and filtering developed by Gordon, Salmond, and Smith (1993) and Kitagawa (1987). Their approach employs discrete fixed-support approximations to unknown densities that appear in the predictive and updating stages of the filtering process. The discrete points that collectively provide density approximations are known as particles; the approach is known as the particle filter. Examples of its use are becoming widespread; in economics, e.g., see Kim, Shephard, and Chib (1998) for an application involving stochastic volatility models; and Fernandez-Villaverde and Rubio-Ramirez (2007) for applications involving dynamic stochastic general equilibrium models.

While conceptually simple and easy to program, the particle filter suffers two shortcomings. First, because the density approximations it provides are discrete, associated likelihood approximations can feature spurious discontinuities, rendering as problematic the application of likelihood maximization procedures (e.g., see Pitt, 2002). Second, the supports upon which

approximations are based are not adapted: period-t approximations are based on supports that incorporate information conveyed by values of the observable variables available in period $t-1$, but not period t (e.g., see Pitt and Shephard, 1999). This gives rise to numerical inefficiencies that can be acute when observable variables are highly informative with regard to state variables, particularly given the presence of outliers.

Numerous extensions of the particle filter have been proposed in attempts to address these problems. For examples, see Pitt and Shephard (1999); the collection of papers in Doucet, de Freitas, and Gordon (2001); Pitt (2002); and the collection housed at http://www-sigproc.eng.cam.ac.uk/ smc/papers.html. Typically, efficiency gains are sought through attempts at adapting period-t densities via the use of information available through period t. However, with the exception of the extension proposed by Pitt (2002), once period-t supports are established they remain fixed over a discrete collection of points as the filter advances forward through the sample, thus failing to address the problem of spurious likelihood discontinuity. (Pitt employs a bootstrap-smoothing approximation designed to address this problem for the specialized case in which the state space is unidimensional.) Moreover, as far as we are aware, no existing extension pursues adaption in a manner that is designed to achieve optimal efficiency.

Here we propose an extension that constructs adapted period-t approximations, but that features a unique combination of two characteristics. The approximations are continuous; and period-t supports are adjusted using a method designed to produce approximations that achieve near-optimal efficiency at the adaption stage. The approximations are constructed using the efficient importance sampling (EIS) methodology developed by Richard and Zhang (2007) (henceforth RZ). Construction is facilitated using an optimization procedure designed to minimize numerical standard errors associated with the approximated integral.

2.2 OVERVIEW OF THE FILTERING PROBLEM

Let y_t be a $n \times 1$ vector of observable variables, and denote $\{y_j\}_{j=1}^t$ as Y_t. Likewise, let s_t be a $m \times 1$ vector of unobserved ('latent') state variables, and denote $\{s_j\}_{j=1}^t$ as S_t . The objective of filtering is to infer the behavior of s_t given Y_t, and an assumed state-space representation; likelihood evaluation obtains as a by-product of the filtering process.

State-space representations consist of a state-transition equation

$$s_t = \gamma(s_{t-1}, Y_{t-1}, \upsilon_t), \tag{2.1}$$

where υ_t is a vector of innovations with respect to (s_{t-1}, Y_{t-1}), and an observation (or measurement) equation

$$y_t = \delta(s_t, Y_{t-1}, u_t), \tag{2.2}$$

where u_t is a vector innovations with respect to (s_t, Y_{t-1}). Hereafter, we refer to υ_t as structural shocks, and u_t as measurement errors.

Filtering is facilitated by interpreting (2.1) and (2.2) in terms of the densities $f(s_t|s_{t-1}, Y_{t-1})$ and $f(y_t|s_t, Y_{t-1})$, respectively. The process is initialized with a marginal density $f(s_0)$, which can be degenerate as a special case. From these densities, the goal is to construct $f(s_t|Y_t)$, which can then be used to calculate, e.g., $E_t(s_t|Y_t)$.

From Bayes' theorem, $f(s_t|Y_t)$ is given by

$$f(s_t|Y_t) = \frac{f(y_t, s_t|Y_{t-1})}{f(y_t|Y_{t-1})} = \frac{f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1})}{f(y_t|Y_{t-1})}, \tag{2.3}$$

where $f(s_t|Y_{t-1})$ is given by

$$f(s_t|Y_{t-1}) = \int f(s_t|s_{t-1}, Y_{t-1}) f(s_{t-1}|Y_{t-1}) ds_{t-1}, \tag{2.4}$$

and $f(y_t|Y_{t-1})$ is given by

$$f(y_t|Y_{t-1}) = \int f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1}) ds_t. \tag{2.5}$$

Note that the recursive structure of $f(s_t|Y_t)$ evident in (2.3) and (2.4) indicates that filtering can be implemented via forward recursion, beginning with the known density $f(s_0) \equiv f(s_0|Y_0)$. Note also that since the likelihood function $f(Y_T)$ factors sequentially as

$$f(Y_T) = \prod_{t=1}^{T} f(y_t|Y_{t-1}), \tag{2.6}$$

where $f(y_1|Y_0) \equiv f(y_1)$, likelihood evaluation obtains as a by-product of the filtering process.

In turn, filtering entails the approximation of the conditional (upon Y_t) expectation of some function $h(s_t)$ (including s_t itself). In light of (2.3) and (2.5), this can be written as

$$E_t(h(s_t)|Y_t) = \frac{\int h(s_t) f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1}) ds_t}{\int f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1}) ds_t}. \tag{2.7}$$

2.3 THE PARTICLE FILTER AND LEADING EXTENSIONS

Since our procedure is an extension of the particle filter, we provide a brief overview here. The particle filter is an algorithm that recursively generates random numbers approximately distributed as $f(s_t|Y_t)$. To characterize its implementation, let $s_t^{r,i}$ denote the i^{th} draw of s_t obtained from the conditional density $f(s_t|Y_{t-r})$ for $r = 0, 1$. A single draw $s_t^{r,i}$ is a particle, and a set of draws $\{s_t^{r,i}\}_{i=1}^N$ is a swarm of particles. The object of filtration is that of transforming a swarm $\{s_{t-1}^{0,i}\}_{i=1}^N$ to $\{s_t^{0,i}\}_{i=1}^N$. The filter is initialized by a swarm $\{s_0^{0,i}\}_{i=1}^N$ drawn from $f(s_0|Y_0) \equiv f(s_0)$.

Period-t filtration takes as input a swarm $\{s_{t-1}^{0,i}\}_{i=1}^N$. The predictive step consists of transforming this swarm into a second swarm $\{s_t^{1,i}\}_{i=1}^N$ according to (2.4). This is done by drawing $s_t^{1,i}$ from the conditional density $f\left(s_t|s_{t-1}^{0,i}, Y_{t-1}\right)$, $i = 1, ..., N$. Note that $\{s_t^{1,i}\}_{i=1}^N$ can be used to produce an MC estimate of $f(y_t|Y_{t-1})$, which according to (2.5) is given by

$$\widehat{f}_N(y_t|Y_{t-1}) = \frac{1}{N}\sum_{i=1}^{N} f(y_t|s_t^{1,i}, Y_{t-1}). \tag{2.8}$$

Next, $f(s_t|Y_t)$ is approximated by re-weighting $\{s_t^{1,i}\}_{i=1}^N$ in accordance with (2.3) (the updating step): a particle $s_t^{1,i}$ with prior weight $\frac{1}{N}$ is assigned the posterior weight

$$w_t^{0,i} = \frac{f(y_t|s_t^{1,i}, Y_{t-1})}{\sum_{j=1}^{N} f(y_t|s_t^{1,j}, Y_{t-1})}. \tag{2.9}$$

The filtered swarm $\{s_t^{0,i}\}_{i=1}^N$ is then obtained by drawing with replacement from the swarm $\{s_t^{1,i}\}_{i=1}^N$ with probabilities $\{w_t^{0,i}\}_{i=1}^N$ (i.e., bootstrapping).

Having characterized the particle filter, its weaknesses (well documented in previous studies) can be pinpointed. First, it provides discrete approximations of $f(s_t|Y_{t-1})$ and $f(s_t|Y_t)$, which moreover are discontinuous functions of the model parameters. The associated likelihood approximation is therefore also discontinuous, rendering the application of maximization routines problematic (a point raised previously, e.g., by Pitt, 2002).

Second, as the filter enters period t, the discrete approximation of $f(s_{t-1}|Y_{t-1})$ is set. Hence the swarm $\{s_t^{1,i}\}_{i=1}^N$ produced in the augmentation stage ignores information provided by y_t. (Pitt and Shephard, 1999 refer to these augmenting draws as "blind".) It follows that if $f(y_t|s_t, Y_{t-1})$ - treated as a function of s_t given Y_t - is sharply peaked in the tails of $f(s_t|Y_{t-1})$, $\{s_t^{1,i}\}_{i=1}^N$ will contain few elements in the relevant range of $f(y_t|s_t, Y_{t-1})$. Thus $\{s_t^{1,i}\}_{i=1}^N$ represents draws from an inefficient sampler: relatively few of its elements will be assigned appreciable weight in the updating stage in the following period. This is known as "sample impoverishment": it entails a reduction in the effective size of the particle swarm.

Extensions of the particle filter employ adaption techniques to generate gains in efficiency. An extension proposed by Gordon, Salmond, and Smith (1993) and Kitagawa (1987) consists simply of making $N' >> N$ blind proposals $\{s_t^{1,j}\}_{j=1}^{N'}$ as with the particle filter, and then obtaining the swarm $\left\{s_t^{0,i}\right\}_{i=1}^{N}$ by sampling with replacement, using weights computed from the N' blind proposals. This is the sampling-importance resampling filter; it seeks to overcome the problem of sample impoverishment by brute force, and can be computationally expensive.

Carpenter and Fernhead (1999) sought to overcome sample impoverishment using a stratified sampling approach to approximate the prediction density. This is accomplished by defining a partition consisting of K subintervals in the state space, and constructing the prediction

density approximation by sampling (with replacement) N_k particles from among the particles in each subinterval. Here N_k is proportional to a weight defined for the entire k^{th} interval; also, $\sum_{k=1}^{K} N_k = N$. This produces wider variation in re-sampled particles, but if the swarm of proposals $\{s_t^{1,i}\}_{i=1}^{N}$ are tightly clustered in the tails of $f(s_t|Y_{t-1})$, so too will be the re-sampled particles.

Pitt and Shephard (1999) developed an extension that ours perhaps most closely resembles. They tackle adaption using an Importance Sampling (IS) procedure. Consider as an example the marginalization step. Faced with the problem of calculating $f(y_t|Y_{t-1})$ in (2.5), but with $f(s_t|Y_{t-1})$ unknown, importance sampling achieves approximation via the introduction into the integral of an importance density $g(s_t|Y_t)$:

$$f(y_t|Y_{t-1}) = \int \frac{f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1})}{g(s_t|Y_t)} g(s_t|Y_t) ds_t. \tag{2.10}$$

Obtaining drawings $s_t^{0,i}$ from $g(s_t|Y_t)$, this integral is approximated as

$$\widehat{f}(y_t|Y_{t-1}) \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f\left(y_t|s_t^{0,i}, Y_{t-1}\right) f\left(s_t^{0,i}|Y_{t-1}\right)}{g(s_t^{0,i}|Y_t)}. \tag{2.11}$$

Pitt and Shephard (1999) referred to the introduction of $g(s_t|Y_t)$ as adaption. Full adaption is achieved when $g(s_t|Y_t)$ is constructed as being proportional to $f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1})$, rendering the ratios in (2.11) as constants. Pitt and Shephard (1999) viewed adaption as computationally infeasible, due to the requirement of computing $f\left(s_t^{0,i}|Y_{t-1}\right)$ for every value of $s_t^{0,i}$ produced by the sampler. Instead they developed samplers designed to yield partial adaption.

The samplers result from Taylor series approximations of $f(y_t|s_t, Y_{t-1})$ around $s_t = \mu_t^k = E\left(s_t|s_{t-1}^{0,k}, Y_{t-1}\right)$. A zero-order expansion yields their auxiliary particle filter; a first-order expansion yields their adapted particle filter. (Smith and Santos, 2006, study examples under which it is possible to construct samplers using second-order expansions.)

These samplers help alleviate blind sampling by reweighting $\left\{s_{t-1}^{0,i}\right\}$ to account for information conveyed by y_t. However, sample impoverishment can remain an issue, since the algorithm does not allow adjustment of the support of $\left\{s_{t-1}^{0,i}\right\}$. Moreover, the samplers are

suboptimal, since μ_t^k is incapable of fully capturing the characteristics of $f(y_t|s_t,Y_{t-1})$. Finally, these samplers remain prone to the discontinuity problem.

Pitt (2002) addressed the discontinuity problem for the special case in which the state space is unidimensional by replacing the weights in (2.9) associated with the particle filter (or comparable weights associated with the auxiliary particle filter) with smoothed versions constructed via a piecewise linear approximation of the empirical c.d.f. associated with the swarm $\left\{s_t^{0,i}\right\}_{i=1}^N$. This enables the use of common random numbers (CRNs) to produce likelihood estimates that are continuous functions of model parameters Hendry (1994).

2.4 THE EIS FILTER

EIS is an automated procedure for constructing continuous importance samplers fully adapted as global approximations to targeted integrands. Section 4.1 outlines the general principle behind EIS, in the context of evaluating (2.5). Section 4.2 introduces a class of piecewise-continuous samplers for dealing with pathological cases. Section 4.3 then discusses a key contribution of this paper: the computation of $f(s_t|Y_{t-1})$ in (2.5) at auxiliary values of s_t generated under period-t EIS optimization. Section 4.4 discusses two special cases that often characterize state-space representations: partial measurement of the state space; and degenerate transition densities. Pseudo-code is available at `www.pitt.edu/~dejong/wp.htm`.

2.4.1 EIS integration

Let $\varphi_t(s_t) = f(y_t|s_t,Y_{t-1})f(s_t|Y_{t-1})$ in (2.5), where the subscript t in φ_t replaces (y_t,Y_{t-1}). Implementation of EIS begins with the preselection of a parametric class $K = \{k(s_t;a_t); a_t \in A\}$ of auxiliary density kernels. Corresponding density functions g are

$$g(s_t;a_t) = \frac{k(s_t;a_t)}{\chi(a_t)},\ \chi(a_t) = \int k(s_t;a_t)ds_t. \tag{2.12}$$

The selection of K is problem-specific; below we discuss Gaussian and piecewise-continuous alternatives. The objective of EIS is to select the parameter value $\widehat{a}_t \in A$ that minimizes the

variance of the ratio $\frac{\varphi_t(s_t)}{g(s_t|a_t)}$ over the range of integration. A (near) optimal value $\widehat{a}_t$ is obtained as the solution to

$$(\widehat{a}_t, \widehat{c}_t) = \arg\min_{a_t, c_t} \int \left[\ln \varphi_t(s_t) - c_t - \ln k(s_t; a_t)\right]^2 g(s_t; a_t) ds_t, \tag{2.13}$$

where c_t is an intercept meant to calibrate $\ln(\varphi_t/k)$. Equation (2.13) is a standard least squares problem, except that the auxiliary sampling density itself depends upon a_t. This is resolved by reinterpreting (2.13) as the search for a fixed-point solution. An operational MC version implemented (typically) using $R << N$ draws, is as follows:

Step $l+1$: Given $\widehat{a}_t^l$, draw intermediate values $\{s_{t,l}^i\}_{i=1}^R$ from the step-l EIS sampler $g(s_t; \widehat{a}_t^l)$, and solve

$$(\widehat{a}_t^{l+1}, \widehat{c}_t^{l+1}) = \arg\min_{a_t, c_t} \sum_{i=1}^{R} \left[\ln \varphi_t(s_{t,l}^i) - c_t - \ln k(s_{t,l}^i; a_t)\right]^2. \tag{2.14}$$

The initial value $\widehat{a}_t^1$ can be chosen in a variety of ways, with minimal impact on convergence. To avoid potential problems involving sample impoverishment, we employ a crude grid search to locate the mode of $\varphi_t(s_t)$. If K belongs to the exponential family of distributions, there exists a parameterization a_t such that the auxiliary problems in (2.14) are linear. Convergence to a fixed point is typically achieved within five to ten iterations.

To guarantee fast fixed-point convergence, and to ensure continuity of corresponding likelihood estimates, $\{s_{t,j}^i\}$ must be obtained by a transformation of a set of common random numbers (CRNs) $\{u_t^i\}$ drawn from a canonical distribution (i.e., one that does not depend on a_t). Examples are standardized Normal draws when g is Gaussian, or uniform draws transformed into draws from g by the inverse c.d.f technique (e.g., see Devroye, 1986).

At convergence, the EIS filter approximation of $f(y_t|Y_{t-1})$ in (2.5) is given by

$$\widehat{f}_N(y_t|Y_{t-1}) = \frac{1}{N} \sum_{i=1}^{N} \frac{f\left(y_t|s_t^i, Y_{t-1}\right) f\left(s_t^i|Y_{t-1}\right)}{g(s_t^i; \widehat{a}_t)}, \tag{2.15}$$

where $\left\{s_t^i\right\}_{i=1}^N$ are drawn from the (final) EIS sampler $g(s_t; \widehat{a}_t)$. This estimate converges almost surely towards $f(y_t|Y_{t-1})$ under weak regularity conditions (outlined, e.g., by Geweke, 1989). Violations of these conditions typically result from the use of samplers with thinner tails than

those of φ_t. RZ offer a diagnostic measure that is adept at detecting this problem. The measure compares the MC sampling variances of the ratio $\frac{\varphi_t}{g}$ under two values of a_t: the optimal $\widehat{a}_t$, and one that inflates the variance of the s_t draws by a factor of 3 to 5.

2.4.2 A piecewise-continuous class of samplers

While kernels within the exponential family of distributions yield EIS regressions that are linear in a_t, there exist potential pathologies of the integrand in (2.5) that they cannot replicate efficiently (e.g., skewness, thick tails, and bimodality). Here we propose an approach that provides high flexibility along one or two pathological dimensions, and as illustrated in Example 4 below, can be combined with (conditional) Gaussian samplers along additional better-behaved dimensions. It entails the use of samplers that provide piecewise log-linear approximations to the integrand φ_t; their parameters are the grid points $a' = (a_0, ..., a_R)$, with $a_0 < a_1 < ... < a_R$ (the index t is suppressed for ease of notation). As we shall see, $\ln k(.;a)$ then depends non-linearly on a. Furthermore, R must be sufficiently large for good approximation. This prevents application of the least-squares optimization step (2.14). Instead we implement near equal probability division of the domain of integration.

We first describe $k(s;a)$ for a preassigned grid a, where the interval $[a_0, a_R]$ covers the support of $\varphi(s)$. Note that while $R+1$ represents the number of grid-points here, and R the number of auxiliary draws used to construct $g(s_t;\widehat{a}_t^l)$ in (2.14), this does not represent an abuse of notation: for the piecewise-continuous sampler, use of $R+1$ grid-points translates precisely into the use of R auxiliary draws.

The kernel $k(s;a)$ is given by

$$\ln k_j(s;a) = \alpha_j + \beta_j s \quad \forall s \in [a_{j-1}, a_j], \tag{2.16}$$

$$\beta_j = \frac{\ln \varphi(a_j) - \ln \varphi(a_{j-1})}{a_j - a_{j-1}}, \quad \alpha_j = \ln \varphi(a_j) - \beta_j a_j. \tag{2.17}$$

Since k is piecewise integrable, its distribution function can be written as

$$K_j(s;a) = \frac{\chi_j(s;a)}{\chi_n(a)}, \quad \forall s \in [a_{j-1}, a_j], \tag{2.18}$$

$$\chi_j(s;a) = \chi_{j-1}(a) + \frac{1}{\beta_j}\left[k_j(s;a) - k_j(a_{j-1};a)\right], \tag{2.19}$$

$$\chi_0(a) = 0, \quad \chi_j(a) = \chi_j(a_j;a). \tag{2.20}$$

Its inverse c.d.f. is given by

$$s = \frac{1}{\beta_j}\left\{\ln\left[k_j(a_{j-1};a) + \beta_j\left(u \cdot \chi_R(a) - \chi_{j-1}(a)\right)\right] - \alpha_j\right\}, \tag{2.21}$$

$$u \in \left]0,1\right[\text{ and } \chi_{j-1}(a) < u \cdot \chi_R(a) < \chi_j(a). \tag{2.22}$$

The recursive construction of an equal-probability-division kernel $k(s;\widehat{a})$ is based upon the non-random equal division of $[\varepsilon, 1-\varepsilon]$ with $u_i = \varepsilon + (2-\varepsilon)\frac{i}{R}$ for $i = 1,...,R-1$, with ε sufficiently small (typically $\varepsilon = 10^{-4}$) to avoid tail intervals of excessive length. It proceeds as follows.

Step $l+1$: Given the step-l grid $\widehat{a}^l$, construct the density kernel k and its c.d.f K as described above. The step-$l+1$ grid is then computed as

$$\widehat{a}_i^{l+1} = K^{-1}(u_i), \quad i = 1,...,R-1. \tag{2.23}$$

The algorithm iterates until (approximate) convergence.

The resulting approximation is highly adapted and computationally inexpensive. Given a sufficiently large number of division points, it will outperform lower-dimensional parametric classes of samplers. Piecewise-continuous samplers can be generalized to higher-dimensional state spaces, though the curse of dimensionality can rapidly become acute. Thus in working with multi-dimensional state spaces, it is advisable to begin with parametric families of distributions, and reserve the use of log-linear piecewise continuous approximations for those dimensions along which the integrand is ill-behaved.

2.4.3 Continuous approximations of $f(s_t|Y_{t-1})$

As noted, the EIS filter requires the evaluation of $f(s_t|Y_{t-1})$ at any value of s_t needed for EIS iterations. Here we discuss three operational alternatives for overcoming this hurdle (a fourth, involving non-parametric approximations, is also possible but omitted here). Below, S denotes the number of points used for each individual evaluation of $f(s_t|Y_{t-1})$.

Weighted-sum approximations

Combining (2.4) and (2.3), we can rewrite $f(s_t|Y_{t-1})$ as a ratio of integrals:

$$f(s_t|Y_{t-1}) = \frac{\int f(s_t|s_{t-1},Y_{t-1})f(y_{t-1}|s_{t-1},Y_{t-2})f(s_{t-1}|Y_{t-2})ds_{t-1}}{\int f(y_{t-1}|s_{t-1},Y_{t-2})f(s_{t-1}|Y_{t-2})ds_{t-1}}, \tag{2.24}$$

where the denominator represents the likelihood integral for which an EIS sampler has been constructed in period $t-1$. A direct MC estimate of $f(s_t|Y_{t-1})$ is given by

$$\widehat{f}_S(s_t|Y_{t-1}) = \frac{\sum_{i=1}^{S} f(s_t|s_{t-1}^{0,i},Y_{t-1})\cdot\omega(s_{t-1}^{0,i};\widehat{a}_{t-1})}{\sum_{i=1}^{S}\omega(s_{t-1}^{0,i};\widehat{a}_{t-1})}, \tag{2.25}$$

where $\{s_{t-1}^{0,i}\}_{i=1}^{S}$ denotes EIS draws from $g(s_{t-1}|\widehat{a}_{t-1})$, and $\left\{\omega(s_{t-1}^{0,i};\widehat{a}_{t-1})\right\}_{i=1}^{S}$ denotes associated weights (both of which are carried over from period-$t-1$), with

$$\omega(s_{t-1};\widehat{a}_{t-1}) = \frac{f(y_{t-1}|s_{t-1},Y_{t-2})f(s_{t-1}|Y_{t-2})}{g(s_{t-1}|\widehat{a}_{t-1})}. \tag{2.26}$$

Obviously $g(s_{t-1}|\widehat{a}_{t-1})$ is not an EIS sampler for the numerator in (2.24). This can impart a potential loss of numerical accuracy if the MC variance of $f(s_t|s_{t-1},Y_{t-1})$ is large over the support of $g(s_{t-1}|\widehat{a}_{t-1})$. This would be the case if the conditional variance of $s_t|s_{t-1},Y_{t-1}$ were significantly smaller than that of $s_{t-1}|Y_{t-1}$. But the fact that we are using the same set of draws for the numerator and the denominator typically creates positive correlation between their respective MC estimators, thus reducing the variance of their ratio.

A constant weight approximation

When EIS delivers a close global approximation to $f(s_{t-1}|Y_{t-1})$, the weights $\omega(s_{t-1};\widehat{a}_{t-1})$ will be near constants over the range of integration. Replacing these weights by their arithmetic means $\overline{\omega}(\widehat{a}_{t-1})$ in (2.24) and (2.25), we obtain the following simplification:

$$f(s_t|Y_{t-1}) \simeq \int f(s_t|s_{t-1},Y_{t-1}) \cdot g(s_{t-1};\widehat{a}_{t-1})ds_{t-1}. \tag{2.27}$$

This substitution yields rapid implementation if additionally the integral in (2.27) has an analytical solution. This will be the case if, e.g., $f(s_t|s_{t-1},Y_{t-1})$ is a conditional normal density for $s_t|s_{t-1}$, and g is either normal or piecewise continuous as described in Section 4.2. Examples are provided in Section 5. In cases for which we lack an analytical solution, we can use the standard MC approximation

$$\widehat{f}_S(s_t|Y_{t-1}) \simeq \frac{1}{S}\sum_{i=1}^{S} f(s_t|s_{t-1}^{0,i},Y_{t-1}). \tag{2.28}$$

EIS evaluation

Evaluation of $f(s_t|Y_{t-1})$ can sometimes be delicate, including situations prone to sample impoverishment (such as when working with degenerate transitions, discussed below). Under such circumstances, one might consider applying EIS not only to the likelihood integral ("outer EIS"), but also to the evaluation of $f(s_t|Y_{t-1})$ itself ("inner EIS").

While outer EIS is applied only once per period, inner EIS must be applied for every value of s_t generated by the former. Also, application of EIS to (2.4) requires the construction of a continuous approximation to $f(s_{t-1}|Y_{t-1})$. Two obvious candidates are as follows. The first is a non-parametric approximation based upon a swarm $\{s_{t-1}^{0,i}\}_{i=1}^{S}$:

$$\widehat{f}_S(s_{t-1}|Y_{t-1}) = \frac{1}{Sh}\sum_{i=1}^{S} \kappa\left(\frac{s_{t-1}-s_{t-1}^{0,i}}{h}\right).$$

The second is the period-$(t-1)$ EIS sampler $g(s_{t-1};\widehat{a}_{t-1})$, under the implicit assumption that the corresponding weights $\omega(s_{t-1};\widehat{a}_{t-1})$ are near-constant, at least over the range of integration. It is expected that in pathological cases, significant gains in accuracy resulting from inner EIS will far outweigh approximation errors in $f(s_{t-1}|Y_{t-1})$.

2.4.4 Special cases

Partial measurement

Partial measurement refers to cases (e.g., see Examples 2 and 4) in which s_t can be partitioned (possibly after transformation) into $s_t = (p_t, q_t)$, so that

$$f(y_t|s_t, Y_{t-1}) \equiv f(y_t|p_t, Y_{t-1}). \tag{2.29}$$

In this case, likelihood evaluation requires integration only with respect to p_t:

$$f(y_t|Y_{t-1}) = \int f(y_t|p_t, Y_{t-1}) f(p_t|Y_{t-1}) dp_t, \tag{2.30}$$

and the updating equation (2.3) factorizes into the product of the following two densities:

$$f(p_t|Y_t) = \frac{f(y_t|p_t, Y_{t-1}) f(p_t|Y_{t-1})}{f(y_t|Y_{t-1})}; \tag{2.31}$$

$$f(q_t|p_t, Y_t) = f(q_t|p_t, Y_{t-1}). \tag{2.32}$$

Stronger conditional independence assumptions are required in order to produce factorizations in (2.4). In particular, if p_t is independent of q_t given (p_{t-1}, Y_{t-1}), so that

$$f(p_t|s_{t-1}, Y_{t-1}) \equiv f(p_t|p_{t-1}, Y_{t-1}), \tag{2.33}$$

then

$$f(p_t|Y_{t-1}) = \int f(p_t|p_{t-1}, Y_{t-1}) f(p_{t-1}|Y_{t-1}) dp_{t-1}. \tag{2.34}$$

Note that under conditions (2.29) and (2.33), likelihood evaluation does not require processing sample information on $\{q_t\}$. The latter is required only if inference on $\{q_t\}$ is itself of interest.

Degenerate transitions

When state transition equations include identities, corresponding transition densities are degenerate (or Dirac) in some of their components. This situation requires an adjustment

to EIS implementation. Again, let s_t partition into $s_t = (p_t, q_t)$, and assume that the transition equations consist of two parts: a proper transition density $f(p_t|s_{t-1}, Y_{t-1})$ for p_t, and an identity for $q_t|p_t, s_{t-1}$ (which could also depend on Y_{t-1}, omitted here for ease of notation):

$$q_t \equiv \phi(p_t, p_{t-1}, q_{t-1}) = \phi(p_t, s_{t-1}). \tag{2.35}$$

The evaluation of $f(s_t|Y_{t-1})$ in (2.4) now requires special attention, since its evaluation at a given s_t (as selected by the EIS algorithm) requires integration in the strict subspace associated with identity (2.35). Note in particular that the presence of identities raises a conditioning issue known as the Borel-Kolmogorov paradox (e.g., see DeGroot, 1984, Section 3.10). We resolve this issue here by reinterpreting (2.35) as the limit of a uniform density for $q_t|p_t, s_{t-1}$ on the interval $[\phi(p_t, s_{t-1}) - \varepsilon, \phi(p_t, s_{t-1}) + \varepsilon]$.

Assuming that $\phi(p_t, s_{t-1})$ is differentiable and strictly monotone in q_{t-1}, with inverse

$$q_{t-1} = \psi(p_t, q_t, p_{t-1}) = \psi(s_t, p_{t-1}) \tag{2.36}$$

and Jacobian

$$J(s_t, p_{t-1}) = \frac{\partial}{\partial q_t}\psi(s_t, p_{t-1}), \tag{2.37}$$

we can take the limit of the integral in (2.35) as ε tends to zero, producing

$$f(s_t|Y_{t-1}) = \int J(s_t, p_{t-1}) f(p_t|s_{t-1}, Y_{t-1}) f(p_{t-1}, q_{t-1}|Y_{t-1})|_{q_{t-1}=\psi(s_t, p_{t-1})} dp_{t-1}. \tag{2.38}$$

Note that (2.38) requires that for any s_t, $f(s_{t-1}|Y_{t-1})$ must be evaluated along the zero-measure subspace $q_{t-1} = \psi(s_t, p_{t-1})$. This rules out use of the weighted-sum approximation introduced above, since the probability that any of the particles $s_{t-1}^{0,i}$ lies in that subspace is zero. We can also approximate the integral in (2.38) by replacing $f(s_{t-1}|Y_{t-1})$ by $\overline{\omega}(\widehat{a}_{t-1}) g(s_{t-1}|\widehat{a}_{t-1})$

$$\widehat{f}(s_t|Y_{t-1}) = \int J(s_t, p_{t-1}) f(p_t|q_{t-1}, Y_{t-1}) g(p_{t-1}, q_{t-1}|\widehat{a}_{t-1})|_{q_{t-1}=\psi(s_t, p_{t-1})} dp_{t-1}. \tag{2.39}$$

In this case, since $g(.|\widehat{a}_{t-1})$ is not a sampler for $p_{t-1}|s_t$, we must evaluate (2.39) either by quadrature or its own EIS sampler.

One might infer from this discussion that the EIS filter is tedious to implement under degenerate transitions, while the particle filter handles such degeneracy trivially in the transition from $\left\{s_{t-1}^{0,i}\right\}$ to $\left\{s_t^{1,i}\right\}$. While this is true, it is also true that these situations are prone to significant sample impoverishment problems, as illustrated in Example 2.

2.5 EXAMPLES

Here we present four examples that illustrate the relative performance of the particle, auxiliary, adapted, and EIS filters. The first two focus on likelihood evaluation; the last two on filtering. We begin with some lessons gleaned through these examples regarding the selection of the three auxiliary sample sizes employed under the EIS filter: N, the number of draws used for likelihood evaluation (e.g., see (2.15)); R , the number of draws used to construct EIS samplers (e.g., see (2.14)); and S, the number of draws used to evaluate $f(s_t|Y_{t-1})$.

First, the efficiency of the EIS filter typically translates into substantial reductions (relative to the particle filter) in the number of draws N needed to reach given levels of numerical accuracy: often by two to three orders of magnitude. In all but the most well-behaved cases, this translates into efficiency gains that more than compensate for the additional calculations required to implement the EIS filter. More importantly, the EIS filter is far more reliable in generating numerically stable and accurate results when confronted with ill-behaved problems (e.g., involving outliers).

Second, in every case we have considered, EIS samplers can be constructed reliably using small values for R (e.g., 100 has sufficed for the applications we have considered).

Third, as with any filter, the range $s_t|Y_{t-1}$ must be sufficiently wide to accommodate period-t surprises (outliers in s_t and/or y_t). At the same time, the approximation grid must be sufficiently fine to accommodate the realization of highly informative realizations of y_t, which generate significant tightening of the distribution of $s_t|Y_t$ relative to that of $s_t|Y_{t-1}$. Both considerations push towards relatively large values for S. The particle filter implicitly sets $N = S$. However, repeated evaluations of $f(s_t|Y_{t-1})$ constitute a substantial portion of EIS computing time, thus setting $S << N$ can yield significant gains in overall efficiency. Indeed, we typically we set $S = 100$. Note that it is trivial to rerun an EIS algorithm under different values for S, thus it is advisable to experiment with alternative values of S in trial runs before launching full-scale analyses in complex applications.

2.5.1 Example 1: Univariate model with frequent outliers

This example is from Fernandez-Villaverde and Rubio-Ramirez (2004b). Let

$$s_{t+1} = \alpha + \beta \frac{s_t}{1+s_t^2} + \upsilon_{t+1} \tag{2.40}$$

$$y_t = s_t + u_t, \tag{2.41}$$

where $\upsilon_{t+1} \sim N\left(0, \sigma_\upsilon^2\right)$ and u_t is t-distributed with ν degrees of freedom:

$$f(u_t) \sim \left(\nu + u_t^2\right)^{-0.5(\nu+1)}, \qquad Var(u_t) = \frac{\nu}{\nu-2} \text{ for } \nu > 2.$$

In all cases, the parameters α and β are both set to 0.5; adjustments to these settings have minimal impact on our results. Note that the expectation of $s_{t+1}|s_t$ is highly non-linear around $s_t = 0$, and becomes virtually constant for $|s_t| > 10$.

We consider two values for ν: 2 and 50. For $\nu = 2$, the variance of u_t is infinite and the model generates frequent outliers: e.g., $\Pr(|u_t| > 4.303) = 0.05$. For $\nu = 50$, u_t is virtually Gaussian: its variance is 1.042, and $\Pr(|u_t| > 2.010) = 0.05$. We consider four values for σ_υ: $(1/3, 1, 3, 10)$. Thus the parameterizations we consider cover a wide range of scenarios, ranging from well-behaved $(\nu = 50, \sigma_\upsilon = 1/3)$ to ill-behaved $(\nu = 2, \sigma_\upsilon = 10)$.

We compare the relative numerical efficiency of five algorithms. The first three are the particle, auxiliary, and adapted filters. These are implemented using $N = 20,000$ and $N = 200,000$. The remaining algorithms are the Gaussian and piecewise-linear EIS filters. These are implemented using $N = 100$ and $N = 1,000$. Evaluation of $f(s_t|Y_{t-1})$ is based on the weighted-sum approximation introduced in Section 4.3.1 – see (2.25).

Results obtained using artificial data sets of size $T = 100$ are presented in Tables 2.1 $(\nu = 2)$ and 2.2 $(\nu = 50)$. Numerical accuracy is assessed by running 100 i.i.d. likelihood evaluations obtained under different seeds. Means of these likelihood evaluations are interpreted as 'final' likelihood estimates; MC standard deviations of these means provide a direct measure of the stochastic numerical accuracy of the final likelihood estimates.

Table 2.1 reports relationships between MC standard deviations and computing time, along with MC means of likelihood evaluations. The tables report a convenient measure of the relative time efficiency of filters i and j:

$$RTE_{i,j} = \frac{T_i V_i}{T_j V_j},$$

where T_i represents computing time per function evaluation, and V_i the MC variance associated with filter i. In the tables, i represents the particle filter; for ratios less than one, the particle filter is the relatively efficient estimator. Reported ratios are based on $N = 200,000$ for the particle, auxiliary and adapted filters, and $N = 1,000$ for the EIS filters.

Note first that RTEs obtained for the auxiliary particle filter range from 0.7 to 1.1 across all cases considered. Thus roughly speaking, regardless of whether the model is well- or ill-behaved, the efficiency gains it generates are offset by associated increases in required computing time, which are on the order of 40%.

Next, for well-behaved cases, RTEs of the adapted particle filter are good; e.g., for $\sigma_\upsilon = 1/3$, efficiency ratios are 8.2 for $\nu = 2$ and 11.6 for $\nu = 50$. However, its performance deteriorates dramatically as σ_υ increases. Indeed, results are not reported for $(\nu = 2, \sigma_\upsilon = 10;$ $\nu = 50, \sigma_\upsilon = 3;)$ and $(\nu = 50, \sigma_\upsilon = 10)$, since in these cases estimated likelihood values diverge pathologically. This reflects the general inability of *local* approximations to provide reliable global approximations of $f(y_t|s_t, Y_{t-1})$ when relevant ranges for s_t become too large. In the present case, problems become critical for Taylor expansions around inflection points of the non-log-concave Student-t density ($y_t = \mu_t^k \pm \sqrt{\nu}$). Note that these are precisely points where second derivatives with respect to s_t are zero, which implies that the use of second-order approximations (e.g., as advocated by Smith and Santos, 2006) would fail to provide an effective remedy in this application.

As expected, RTEs of the Gaussian EIS filter are also poor given $\nu = 2$, especially when σ_υ is large. For $\nu = 50$, the Gaussian EIS filter performs well, with impressive RTEs for large values of σ_υ (reaching 284 for $\sigma_\upsilon = 10$).

The piecewise-linear EIS filter outperforms the particle filter in all cases, with the payoff to adoption increasing with σ_υ. For $\nu = 2$, its RTE ranges from 1.6 to 2,001 as σ_υ increases

from 1/3 to 10; for $\nu = 50$, its RTE ranges from 2.3 to 1,401. An RTE of 1,400 implies that the particle filter requires approximately 1 hour and 15 minutes (the time required to process approximately 38.4 million particles) to match the numerical accuracy of the piecewise-linear filter with $N = 1,000$ (which requires 3.18 seconds). These results reflect the payoffs associated with the flexibility, in addition to the global nature, of approximations provided by the piecewise-linear filter.

In sum, the particle, auxiliary, and adapted filters perform well under well-behaved scenarios. In these cases, their relative numerical inaccuracy is often offset by their relative speed. However, expansions in the range of s_t, along with the presence of outliers, can lead to dramatic reductions in RTEs, and in the case of the auxiliary and adapted filters, can also lead to unreliable likelihood estimates. The EIS filters provide insurance against these problems and exhibit superior RTEs in all but the most well-behaved cases. But while numerical efficiency is an important feature of likelihood approximation procedures, it is not the only important feature. In pursuing ML estimates, continuity with respect to parameters is also critical. The next example highlights this feature.

2.5.2 Example 2: A dynamic stochastic general equilibrium model

Following Sargent (1989), likelihood-based analyses of dynamic stochastic general equilibrium (DSGE) models have long involved the application of the Kalman filter to log-linear model approximations (e.g., see DeJong, Ingram, and Whiteman, 2000; Otrok, 2001; Ireland, 2004; and the survey by An and Schorfheide, 2007). However, Fernandez-Villaverde, Rubio-Ramirez, and Santos (2006) have shown that second-order approximation errors in model solutions map into first-order effects on the corresponding likelihood function, due to the accumulation over time of approximation errors. Fernandez-Villaverde and Rubio-Ramirez (2005) document the quantitative relevance of this phenomenon in an empirical analysis involving estimates of a neoclassical growth model obtained using the particle filter.

Here we demonstrate the performance of the EIS filter by estimating the structural parameters of a simple growth model via maximum likelihood. Regarding the model, let q_t, k_t, c_t, i_t,

and a_t represent output, capital, consumption, investment, and total factor productivity (TFP). Labor is supplied inelastically and fixed at unity. The model is of a representative agent who seeks to maximize the expected value of lifetime utility

$$U = E_0 \sum_{t=0}^{\infty} \beta^t \ln(c_t),$$

subject to

$$\begin{aligned} q_t &= a_t k_t^{\alpha} && (2.42)\\ q_t &= c_t + i_t && (2.43)\\ k_{t+1} &= i_t + (1-\delta)k_t && (2.44)\\ \ln(a_{t+1}) &= \rho \ln(a_t) + \upsilon_{t+1}. && (2.45) \end{aligned}$$

Regarding parameters, α is capital's share of output, δ is capital depreciation, ρ determines the persistence of innovations to TFP, and the innovations $\upsilon_t \sim N(0, \sigma^2)$. The state variables (a_t, k_t) are unobserved, and the distribution of (a_0, k_0) is known. The solution of this problem can be represented as a policy function for consumption of the form $c(a_t, k_t)$. For the special case in which $\delta = 1$, $c(a_t, k_t) = (1 - \alpha\beta) a_t k_t^{\alpha}$. This is the case studied here.

We take q_t and i_t as observable, subject to measurement error. Combining equations, the measurement equations are

$$\begin{aligned} q_t &= a_t k_t^{\alpha} + u_{q_t}, && (2.46)\\ i_t &= a_t k_t^{\alpha} - c(a_t, k_t) + u_{i_t} && (2.47)\\ &= \alpha\beta a_t k_t^{\alpha} + u_{i_t}, \end{aligned}$$

and the state-transition equations are (2.45) and

$$\begin{aligned} k_{t+1} &= a_t k_t^{\alpha} - c(a_t, k_t) && (2.48)\\ &= \alpha\beta a_t k_t^{\alpha}. \end{aligned}$$

Examination of (2.45) to (2.48) suggests reparameterizing the state variables as $z_t = \ln(a_t)$ and $l_t = e^{z_t} k_t^{\alpha}$, where l_t denotes (unobserved) output, and $s_t = [l_t \quad z_t]'$ denotes the state vector.

The transition process (2.45) then takes the form of a Gaussian AR(1) in z_t, and the identity (2.48) can be rewritten as

$$l_t = e^{z_t} \left(\alpha\beta l_{t-1}\right)^{\alpha}. \tag{2.49}$$

Note that this example combines the two special cases discussed in Section 4.4. First, there is partial measurement, in that y_t is independent of z_t conditionally on l_t (and Y_{t-1}):

$$y_t|s_t, Y_{t-1} \equiv y_t|l_t, Y_{t-1} \sim N_2\left(\begin{pmatrix} 1 \\ \alpha\beta \end{pmatrix} l_t, \begin{bmatrix} \sigma_q^2 & 0 \\ 0 & \sigma_i^2 \end{bmatrix}\right). \tag{2.50}$$

Second, (2.49) represents a degenerate Dirac transition, with inverse

$$l_{t-1} = \psi(s_t) = \frac{1}{\alpha\beta}\left(l_t e^{-z_t}\right)^{\frac{1}{\alpha}} \tag{2.51}$$

and Jacobian

$$J(s_t) = \frac{\partial\psi(s_t)}{\partial l_t} = \frac{1}{\alpha^2\beta}\left(l_t^{1-\alpha} e^{-z_t}\right)^{\frac{1}{\alpha}}. \tag{2.52}$$

In view of (2.50), the likelihood integral simplifies into a univariate integral in l_t whose evaluation requires only an EIS sampler for $l_t|Y_t$. Nevertheless, in period $t+1$ we still need to approximate $f(z_t|l_t, Y_t)$ in order to compute $\widehat{f}(y_t|Y_{t-1})$. To capture the dependence between z_t and l_t given Y_t, it proves convenient to construct a single bivariate EIS sampler for $z_t, l_t|Y_t$. Whence the likelihood integral

$$f(y_t|Y_{t-1}) = \int f(y_t|l_t, Y_{t-1}) f(s_t|Y_{t-1}) ds_t \tag{2.53}$$

is evaluated under a Gaussian EIS sampler $g(s_t|\hat{a}_t)$. Next, $f(s_t|Y_{t-1})$ is approximated according to (2.39), where we exploit the fact that the Jacobian $J(s_t)$ does not depend on z_{t-1}:

$$\widehat{f}(s_t|Y_{t-1}) = J(s_t) \int f(z_t|z_{t-1}) g(\psi(s_t), z_{t-1}|\hat{a}_{t-1}) dz_{t-1}. \tag{2.54}$$

Note that the integrand is quadratic in $z_{t-1}|s_t$, so standard algebraic operations amounting to the completion of a quadratic form in z_{t-1} yield an analytical solution for $\widehat{f}(s_t|Y_{t-1})$. Thus under the implicit assumption that the EIS weights $\omega(s_t; \hat{a}_t)$ are near constant (to be verified empirically), we have derived a particularly fast and efficient EIS implementation based on a bivariate Gaussian outer EIS, and an inner analytical approximation for $f(s_t|Y_{t-1})$.

Model estimates are based on artificial data simulated from the model. Parameter values used to simulate the data are as follows: $\alpha = 0.33$, $\beta = 0.96$, $\rho = 0.8$, $\sigma = 0.05$, $\sigma_q = 0.014$, $\sigma_i = 0.02$. The first four values are typical of this model calibrated to annual data; and given σ, the latter two values represent approximately 5% and 20% of the unconditional standard deviations of q_t and i_t. The unconditional mean and standard deviation of a_t implied by ρ and σ equal 1.0035 and 0.08378.

To begin, we compute likelihood values at the true parameter values using 20,000 and 100,000 particles for the particle filter, and 100 and 1,000 auxiliary draws for the EIS filter (with R held fixed at 100). We do so for 100 MC replications, applied to a single data set of sample size $T = 100$. Results are reported in Table 2.3.

RTEs computed using as a numeraire the particle filter with $N = 20,000$ are 55.44 (for $N = 100$ under the EIS filter) and 217.728 (for $N = 1,000$). That is, the time required for the particle filter to attain the same standard of numerical accuracy exceeds the time required by the EIS filter with $N = 1,000$ by a factor of approximately 217 (the time required to process approximately 8.7 million particles). This difference in efficiency is due to the fact that the bivariate Gaussian EIS samplers $g(s_t|\widehat{a}_t)$ provide close (global) approximations of the densities $f(s_t|Y_{t-1})$. Indeed, on a period-by-period basis, ratios of standard deviations to the means of the weights $\left\{\omega\left(s_t^{0,i};\widehat{a}_t\right)\right\}_{i=1}^{N}$ range from 1.14e-8 to 3.38e-3. Such small variations validate our reliance on (2.54) to approximate $f(s_t|Y_{t-1})$.

Next, we apply both the particle and EIS filters to compute maximum likelihood estimates (MLEs) for $\theta = (\alpha, \beta, \rho, \sigma, \sigma_q, \sigma_i)$, under simulated samples of size T=40, 100 and 500. Using (2.47), the stepwise MLE of β given α is given by $\widehat{\beta} = \bar{i}/\alpha\bar{l}$, where $\bar{i}$ and $\bar{l}$ denote sample means of i_t and l_t. MLEs for the remaining parameters are obtained via maximization of the concentrated log-likelihood function. Results for the particle filter are based on $N = 20,000$; results for the EIS filter are based on $N = 200$ and $R = 100$. Note from Table 2.3 that computing times for a single likelihood evaluation are approximately the same under both methods (on the order of 5.5 seconds for $T = 100$), while MC estimates of the log-likelihood function are much more accurate under the EIS filter (which has an RTE of approximately 55 given

these settings for N and R).

Figure 2.1 plots estimated log-likelihoods for a representative data set along the α dimension for $T = 100$ for both the particle and EIS filters; all other parameters are set at their ML values. Note that the surface associated with the particle filter is particularly rough relative to that associated with EIS.

We employ the Nelder-Mead simplex optimization routine for all MLE computations. Following RZ, we use i.i.d replications (30 in the present set-up) of the complete ML algorithm in order to produce two sets of means and standard deviations for MLEs. The first are *statistical* means and standard deviations, obtained from 30 different samples $\{y_t\}_{t=1}^T$ under a single set of auxiliary draws $\{u_i\}_{i=1}^N$. These characterize the finite sample distribution of the MLEs. Under the EIS filter, we also compute the asymptotic standard deviations obtained by inversion of a numerical Hessian. As in Fernandez-Villaverde and Rubio-Ramirez (2007), we find that Hessians computed under the particle filter are unreliable and often fail to be positive definite. The second are *numerical* means and standard deviations, obtained under 30 different sets of CRNs for a fixed sample $\{y_t\}_{t=1}^T$. These constitute our most accurate MC estimates of the MLEs and accordingly, the numerical standard deviations we report are those for the means.

Results are given in Table 2.4. Highlights are as follows. (1) Log-likelihood functions are tightly peaked, as statistical standard deviations attest. (2) For $T = 40$, MLEs of α are upward biased (by about 4 standard deviations), thus we also report root mean-squared errors. (3) Under the EIS filter, there is close agreement between finite sample (MC) and asymptotic (Hessian) statistical standard deviations, especially as T increases. This highlights the numerical accuracy and reliability of EIS filter computations (including Hessians). (4) As T increases, numerical standard deviations (which are $\sqrt{30}$ larger than those reported for the mean MLEs) approach corresponding statistical standard deviations. This does not create a problem for the EIS filter (which employs CRNs), but contaminates the computation of statistical standard deviations under the particle filter. For this example, N would need to be increased dramatically in order for the particle filter to provide reliable estimates of statistical standard deviations.

In sum, MLEs derived using the EIS filter ($N = 200$, $R = 100$) are numerically and statistically significantly more reliable than those derived under the particle filter ($N = 20,000$). They are also obtained relatively more rapidly (by a factor of 25% to 50%).

2.5.3 Example 3: Stochastic Volatility

The stochastic volatility (SV) model is given by

$$y_t = u_t \beta \exp(s_t/2) \tag{2.55}$$

$$s_{t+1} = \phi s_t + \upsilon_t, \tag{2.56}$$

where u_t and υ_t are independent Gaussian random variables with variances 1 and σ^2; β represents modal volatility; and ϕ and σ^2 determine the persistence and variance of volatility shocks. This model was introduced by Taylor (1982, 1986) in attempts to account for the time-varying and persistent volatility exhibited by financial returns data, in addition to fat-tailed behavior. Many alternative procedures have been proposed to estimate this model efficiently, and to infer the behavior of (scaled) volatility (e.g., see Jacquier, Polson, and Rossi, 1994; Ghysels, Harvey, and Renault, 1996; Pitt and Shephard, 1999; Kim, Shephard, and Chib, 1998; and Liesenfeld and Richard, 2003). Thus it provides a natural testing ground for us as well.

Filtered values for volatility are obtained by replacing $h(s_t)$ by $\exp(s_t/2)$ in (2.7). These are obtained using Gaussian EIS samplers. Due to ease of implementation in the present context, we construct separate samplers for the numerator and denominator of (2.7).

Consider the approximation of the denominator: the period-t likelihood $f(y_t|Y_{t-1})$. Let $g_t(s_t;a_{d,t})$ denote the EIS sampling density used to approximate the integrand $\varphi_t(s_t) = f(y_t|s_t,Y_{t-1})f(s_t|Y_{t-1})$. The associated log kernel is parameterized as

$$-2\ln k(s_t;a_{d,t}) = \alpha_{d,t}s_t^2 - 2\beta_{d,t}s_t, \tag{2.57}$$

where $a_{d,t} = (\alpha_{d,t},\beta_{d,t})$. The corresponding Gaussian sampler has mean $\mu_{d,t} = \beta_{d,t}/\alpha_{d,t}$ and variance $\sigma^2_{d,t} = \alpha^{-1}_{d,t}$. Thus the auxiliary regression for the computation of $\hat{a}_{d,t}$ consists of a

bivariate OLS regression of simulated values of $\ln \varphi_t(s_t)$ on simulated values of s_t^2, s_t and a constant.

We use a constant-weight approximation to approximate the prediction density $f(s_t|Y_{t-1})$; the approximation obtains as

$$f(s_t|Y_{t-1}) \approx \int f(s_t|s_{t-1},Y_{t-1})g(s_{t-1};\hat{a}_{d,t-1})ds_{t-1}. \tag{2.58}$$

Since $f(s_t|s_{t-1},Y_{t-1})$ is a conditional normal density for $s_t|s_{t-1},Y_{t-1}$, and $g(s_{t-1};\hat{a}_{d,t-1})$ is a normal density for s_{t-1}, the integral in (2.58) has an analytical solution given by a Gaussian density for s_t with mean $\phi\hat{\mu}_{d,t-1}$ and variance $\phi^2\hat{\sigma}^2_{d,t-1}+\sigma^2_{\upsilon}$. Assuming $f(s_1|Y_0)=f(s_1)$, the initial values are $\hat{\mu}_{d,0}=0$ and $\hat{\sigma}^2_{d,0}=\sigma^2_{\upsilon}/(1-\phi^2)$.

Given an EIS sampler for the denominator, the Gaussian EIS kernel for the numerator in (2.7), denoted by $k(s_t;\hat{a}_{n,t})$ and designed to approximate $h(s_t)\cdot\varphi_t(s_t)$, is obtained analytically, since in the present context $\ln h(s_t)$ is a linear function in s_t. In particular, optimal values for the mean and variance of the sampler for the numerator $(\hat{\mu}_{n,t},\hat{\sigma}^2_{n,t})$ are given by $\hat{\sigma}^2_{n,t}=\hat{\sigma}^2_{d,t}$ and $\hat{\mu}_{n,t}=\hat{\sigma}^2_{n,t}(\hat{\mu}_{d,t}/\hat{\sigma}^2_{d,t}+1/2)$. Hence the construction of the optimal sampler for the numerator is obtained without incurring additional computing costs.

Based on these EIS samplers, a consistent MC estimate of the filtered values obtains as

$$\widehat{E}\left(\exp(s_t/2)|Y_t\right)=\frac{\sum_{i=1}^{N} w_{n,t}\left(\tilde{s}^{0,i}_{n,t}\right)}{\sum_{i=1}^{N} w_{d,t}\left(\tilde{s}^{0,i}_{d,t}\right)}, \tag{2.59}$$

where $w_{n,t}(s_t)=\exp(s_t/2)\cdot\varphi_t(s_t)/g(s_t;\hat{a}_{n,t})$ and $w_{d,t}(s_t)=\varphi_t(s_t)/g(s_t;\hat{a}_{d,t})$. The sets $\{\tilde{s}^{0,i}_{n,t}\}$ and $\{\tilde{s}^{0,i}_{d,t}\}$ denote swarms of iid draws from $g(s_t;\hat{a}_{n,t})$ and $g(s_t;\hat{a}_{d,t})$, generated under a single set of CRNs.

We demonstrate the performance of the EIS filter for the SV model in an application to sets of artificial data simulated from the model. Performance is characterized relative to that of the particle filter, using the exact experimental design used by Pitt and Shephard (1999) to characterize the performance of their auxiliary and adapted filters.

The model is parameterized as $\phi = 0.9702$, $\sigma_\upsilon = 0.178$, and $\beta = 0.5992$; the sample size is $T = 50$. We draw $R = 40$ different data sets $\{Y_T^i\}_{i=1}^{40}$, all based on one simulated trajectory of the latent variable $\{s_t\}_{t=1}^{50}$. In the measurement-error series, we artificially insert a single outlier: $u_{21} = 2.5$. For each data set we produce 100 i.i.d. MC estimates of the filtered values for volatility using EIS-1K and PF-20K, where the numbers following the acronyms indicate the number of particles (computing times are similar given these settings). For each procedure, the 100 MC estimates are obtained using 100 different CRNs. Comparing MC estimates generated by these procedures with "true" values of the filtered means yields Mean Squared Error (MSE) comparisons identical to those used by Pitt and Shephard (1999).

Let $\tilde{\ell}_t^i$, $(i : 1 \to 40,\ t : 1 \to 50)$ denote "true" filtered means for volatility. These must be computed with high numerical accuracy in order to validate the MSE comparisons that follow. Exploiting the relatively high numerical accuracy of EIS (highlighted below), we estimate "true" filtered means as the arithmetic means of 100 i.i.d. EIS-10K estimates. Corresponding standard deviations are several orders of magnitude lower than those of the estimates we propose to compare. In order to reach similar precision using the particle filter, we must use the arithmetic means of 100 i.i.d. PF-4 million estimates. We ran this experiment to verify that "true" values produced by both EIS and PF estimators are numerically identical. The latter number far exceeds the PF-120K value employed by Pitt and Shephard (1999), but turns out to be needed in order to eliminate significant and persistent biases characterizing PF estimates of filtered means (illustrated below).

MSE comparisons are constructed as follows. Let $\bar{\ell}_{t,k}^{i,j}$ denote the MC estimate of the filtered mean, for data set i, for replication j, at time t, for procedure $k = \{\text{EIS-1K}, \text{PF-20K}\}$. The log mean squared error (LMSE) for procedure k, at time t is obtained as

$$\text{LMSE}_{t,k} = \ln\left\{\frac{1}{40}\sum_{i=1}^{40}\left[\frac{1}{100}\sum_{j=1}^{100}\left(\bar{\ell}_{t,k}^{i,j} - \tilde{\ell}_t^i\right)^2\right]\right\}. \tag{2.60}$$

Figure 2.2 (bottom panel) depicts LMSEs for the five procedures against time. As expected, the move from estimates obtained using the particle filter to those obtained using the EIS filter leads to a large reduction in LMSEs: the average difference between PF-20K and

EIS-1K is 1.9 on log scale. These differences are far larger than those reported by Pitt and Shephard (1999), (see Figure 4 in Pitt and Shephard, 1999) in their comparison of the particle filter with their adapted particle filter, both implemented using the same number of particles. In particular, their adapted filter yields a maximal reduction of 1.0 (0.8) on log scale relative to the particle filter using 2K (4K) particles. Note also that the EIS filter is considerably less susceptible to the injected outlier in the measurement error process than is the particle filter.

To identify the source of the large differences in LMSEs, we computed separately MC variances and squared biases for EIS-1K and PF-20K. Logged variances and logged MSE/variance ratios are plotted for both procedures in Figure 2.2 (top panels). The logged MSE/variance ratio can be interpreted as a "bias multiplier", indicating the extent to which biases amplify differences in logged variances in yielding corresponding LMSEs. Figure 2.2 indicates that in nearly all periods the logged MC variance for the EIS filter is substantially smaller than for the particle filter. Further, EIS-1K exhibits logged MSE/variance ratios close to zero for all time periods, indicating near-complete absence of bias. In contrast, for PF-20K this ratio is significantly larger than zero in approximately half of the time periods. Note in particular the comparably large value of the ratio for PF-20K in the time period infected by the outlier $(t = 21)$. These results indicate that, in addition to MC variance, bias represents a significant component of the large differences in LMSEs generated by the adoption of EIS.

2.5.4 Example 4: Bearings-Only Tracking

The bearings-only tracking problem has received much attention in the literature on particle filters, and raises challenging numerical issues. References include Gordon, Salmond, and Smith (1993), Carpenter and Fernhead (1999), and Pitt and Shephard (1999); we consider here the scenario described by Gordon, Salmond, and Smith (1993).

A ship moves in the (x,z) plane with speed following a bivariate random walk process. Let $\lambda_t = \left(x_t, \quad z_t, \quad \dot{x}_t, \quad \dot{z}_t\right)'$ denote the quadrivariate latent state variable (shortly we shall re-parameterize, and revert to the use of s_t to denote the state). The discrete version of the

model we consider first is characterized by the transition

$$\lambda_{t+1} = \begin{pmatrix} I_2 & I_2 \\ 0 & I_2 \end{pmatrix} \lambda_t + \sigma \begin{pmatrix} \frac{1}{2} I_2 \\ I_2 \end{pmatrix} u_t, \tag{2.61}$$

with $u_t \sim iidN(0, I_2)$. The initial state vector is distributed as

$$\lambda_1 \sim N(\mu_1, \Delta_1), \tag{2.62}$$

with (μ_1, Δ_1) known and Δ_1 diagonal.

An observer located at the origin of the (x,z) plane measures with error the angle $\theta_t = \arctan(z_t/x_t)$. The measured angle y_t is assumed to be wrapped Cauchy with density

$$f(y_t|\theta_t) = \frac{1}{2\pi} \frac{1-r^2}{1+r^2-2r\cos(y_t-\theta_t)}, \tag{2.63}$$

with $0 \leq (y_t, \theta_t) \leq 2\pi$ and $0 \leq r \leq 1$. Accordingly, we shall introduce a (partial) reparameterization in polar coordinates. Let

$$\begin{aligned} \lambda_t &= (\alpha_t, \quad \beta_t)' \qquad \alpha_t = (x_t, \quad z_t)' \qquad \beta_t = \left(\dot{x}_t, \quad \dot{z}_t\right)', \\ \alpha_t &= \rho_t e(\theta_t), \qquad \theta_t \in [0, 2\pi], \end{aligned}$$

with $e(\theta_t) = (\cos\theta_t, \quad \sin\theta_t)'$ and $\rho_t = \left(x_t^2 + z_t^2\right)^{1/2} \geq 0$. The following notation will be used for the transformed state vector:

$$s_t = h(\lambda_t) = (\theta_t, \quad \rho_t, \quad \beta_t') = (\theta_t, \quad \delta_t'). \tag{2.64}$$

Note that (2.61) is based on a discretization over a time interval that coincides with the interval between successive measurements. It implies that the transition from λ_t to λ_{t+1} is degenerate. We reinterpret this transition as the combination of a proper bivariate transition

$$\alpha_{t+1}|\lambda_t \sim N(A\lambda_t, \Omega) \tag{2.65}$$

and a Dirac transition

$$\beta_{t+1} \equiv \phi(\alpha_{t+1}, \lambda_t) = 2(\alpha_{t+1} - \alpha_t) - \beta_t, \tag{2.66}$$

with $A = (I_2, \quad I_2)$ and $\Omega = \frac{1}{4}\sigma^2 I_2$. Below we shall consider an alternative version of the model discretized on a finer grid than that defined by observation times. This produces a non-degenerate transition, and allows for observations that are spaced unequally over time.

The degenerate version just described is numerically challenging on three counts. First, measurement is non-informative on three out of the four state components. Second, under parameter values typically used in the literature, the density of $\theta_t|Y_t$ is much tighter (though with fat tails) than that of $\theta_t|Y_{t-1}$. This situation yields "sample impoverishment", and thus (very) high numerical inefficiency for the particle filter. Finally, the degenerate transition creates additional numerical problems since it implies a zero-measure support in $\mathbb{R}^4$ for the density $f(\lambda_{t+1}|\lambda_t)$.

Despite these challenges, we can implement an EIS version of the particle filter that can accommodate these pathologies. While conceptually simple, the algebra of our implementation is somewhat tedious. The text presents the broad lines of our implementation; full technical details are regrouped in the Appendix.

EIS computation of $f(y_t|Y_{t-1})$

We momentarily take as given that $f(s_t|Y_{t-1})$ can be computed for any s_t (as described below). The period-t likelihood function is then given by

$$\ell_t \equiv f(y_t|Y_{t-1}) = \int f(y_t|\theta_t) f(s_t|Y_{t-1}) ds_t. \tag{2.67}$$

Note that while $f(s_t|s_{t-1})$ is degenerate, $f(s_t|Y_{t-1})$ is not. In the absence of observations, $f(s_t)$ would be quadrivariate Normal. The observation y_t only measures θ_t, thus we shall implement a (sequential) EIS sampler $g(s_t; a_t)$ as the product of a trivariate Gaussian density for $\delta_t|\theta_t$ and a univariate piecewise loglinear density for θ_t. For ease of notation, the auxiliary EIS parameter a_t is deleted from all subsequent equations.

The conditional EIS sampler $g(\delta_t; \theta_t)$ is constructed as follows (accounting for the transformation from λ_t to s_t): (i) We draw a swarm $\left\{\widetilde{\lambda}_t^{1,i}\right\}_{i=1}^N$. Specifically, the period-$(t-1)$ EIS swarm $\left\{\widetilde{s}_{t-1}^{0,i}\right\}_{i=1}^N$ is transformed into a swarm $\left\{\widetilde{\lambda}_{t-1}^{0,i}\right\}_{i=1}^N$ by means of the inverse transformation $\lambda_t = h^{-1}(s_t)$. Then $\widetilde{\lambda}_t^{1,i}$ is drawn from the (degenerate) transition density $f\left(\lambda_t|\widetilde{\lambda}_{t-1}^{0,i}\right)$ associated with (2.61).

(ii) We construct an auxiliary quadrivariate EIS Gaussian kernel $k_{\lambda,t}(\lambda_t)$ approximating $f(\lambda_t|Y_{t-1})$. To do so, we use the swarm $\left\{\widetilde{\lambda}_t^{1,i}\right\}_{i=1}^N$ to construct an auxiliary OLS regression of $\left\{\ln f\left(\widetilde{\lambda}_t^{1,i}|Y_{t-1}\right)\right\}_{i=1}^N$ on $\left\{\widetilde{\lambda}_t^{1,i}\right\}_{i=1}^N$ and the lower triangle of $\left\{\left(\widetilde{\lambda}_t^{1,i}\right)\left(\widetilde{\lambda}_t^{1,i}\right)'\right\}_{i=1}^N$, for a total of 14 regressors plus one intercept. Let μ_t denote the unconditional mean of this quadrivariate kernel and P_t its precision matrix. The kernel is then written as

$$k_{\lambda,t}(\lambda_t) = \exp\left\{-\frac{1}{2}\left(\lambda_t' P_t \lambda_t - 2\lambda_t' q_t\right)\right\}, \tag{2.68}$$

with $q_t = P_t^{-1}\mu_t$.

(iii) We introduce the transformation from λ_t to $s_t = h(\lambda_t)$, with Jacobian $\rho_t > 0$. Let

$$k_{s,t}(s_t) = \rho_t k_{\lambda,t}\left(h_t^{-1}(s_t)\right). \tag{2.69}$$

The conditional EIS sampler for $\delta_t|\theta_t$ is then given by

$$g_t(\delta_t|\theta_t) = \frac{k_{s,t}(s_t)}{\chi_t(\theta_t)}, \tag{2.70}$$

with

$$\chi_t(\theta_t) = \int_\Delta k_{s,t}(s_t)\, d\beta_t d\rho_t, \tag{2.71}$$

where $\Delta = \mathbb{R}^2 \times \mathbb{R}_+$.

(iv) The likelihood integral in (2.67) is rewritten as

$$\ell_t = \int \left[f(y_t|\theta_t)\chi_t(\theta_t)\right] \frac{f(s_t|Y_{t-1})}{k_{s,t}(s_t)} g_t(\delta_t|\theta_t)\, d\delta_t d\theta_t. \tag{2.72}$$

The next EIS step consists of approximating the product $f(y_t|\theta_t)\chi_t(\theta_t)$ on $[0, 2\pi]$ by a piecewise loglinear EIS sampler $g_t(\theta_t)$. Equation (2.72) is rewritten as

$$\ell_t = \int w_{s,t}(s_t)\, g_t(\delta_t|\theta_t)\, g_t(\theta_t)\, ds_t, \tag{2.73}$$

with

$$w_{s,t}(s_t) = \left[\frac{f(y_t|\theta_t)\chi_t(\theta_t)}{g_t(\theta_t)}\right] \frac{f(s_t|Y_{t-1})}{k_{s,t}(s_t)}. \tag{2.74}$$

Its EIS-MC estimate obtains as

$$\widehat{\ell_t} = \frac{1}{N}\sum_{i=1}^{N} w_{s,t}\left(\widetilde{s}_t^{\theta,i}\right), \tag{2.75}$$

where $\left\{\widetilde{s}_t^{\theta,i}\right\}_{i=1}^{N}$ denotes a swarm of i.i.d.N. draws (under CRNs) from the EIS sampler $g(\delta_t|\theta_t)\, g_t(\theta_t)$.

In view of the structure of the problem (non-observability of 3 out of 4 Gaussian state variables, and flexibility of the piecewise loglinear sampler along the fourth), we anticipate close fit between the numerator and denominator of $w_{s,t}(s_t)$ as given in (2.74). Relatedly, we anticipate dramatic reduction in the MC sampling variance of filtered values relative to that of estimates obtained under the particle filter and commonly used extensions.

EIS computation of $f(\lambda_{t+1}|Y_t)$

Having just discussed EIS for period t, it is notationally more convenient to discuss the computation of $f(\lambda_{t+1}|Y_t)$ rather than that of $f(\lambda_t|Y_{t-1})$. The reason for initially discussing $f(\lambda_{t+1}|Y_t)$ rather than $f(s_{t+1}|Y_t)$ is simply that Gaussian algebraic manipulations are more transparent under the λ parametrization. Moreover, $f(s_{t+1}|Y_t)$ obtains directly from $f(\lambda_{t+1}|Y_t)$ via the transformation $s_{t+1} = h(\lambda_{t+1})$ with Jacobian $\rho_t > 0$. Relatedly, the weights $w_{s,t}(s_t)$ in (2.74) can trivially be transformed into weights for λ_t. Let

$$w_{\lambda,t}(\lambda_t) = w_{s,t}\left(h^{-1}(s_t)\right) = \frac{f(\lambda_t|Y_{t-1})}{k_{\lambda,t}(\lambda_t)}\left[\frac{f(y_t|\theta_t)\,\chi_t(\theta_t)}{g_t(\theta_t)}\right]_{\theta_t=\theta_t(\alpha_t)}, \tag{2.76}$$

with $\theta_t(\alpha_t) = \arctan(z_t/x_t)$. Whence the density $f(\lambda_t|Y_t)$, given by

$$f(\lambda_t|Y_t) = \frac{f(\lambda_t|Y_{t-1})\, f(y_t|\theta_t(\alpha_t))}{\ell_t}, \tag{2.77}$$

can be rewritten as

$$f(\lambda_t|Y_t) = \frac{w_{\lambda,t}(\lambda_t)}{\ell_t} k_{\lambda,t}(\lambda_t)\frac{g_t(\theta_t)}{\chi_t(\theta_t)}|_{\theta_t=\theta_t(\alpha_t)}. \tag{2.78}$$

Under a non degenerate transition from λ_t to λ_{t+1}, $f(\lambda_{t+1}|Y_t)$ obtains as

$$f(\lambda_{t+1}|Y_t) = \int_{\mathbb{R}^4} f(\lambda_t|Y_t)\, f(\lambda_{t+1}|\lambda_t)\, d\lambda_t. \tag{2.79}$$

In the present case, however, we have to properly account for the fact that the transition from λ_t to λ_{t+1} is degenerate. As discussed above, degeneracy is addressed by replacing β_t in (2.78) by the inverse of the Dirac transition in (2.66):

$$\phi^{-1}(\lambda_{t+1}, \alpha_t) = \beta_t = 2(\alpha_{t+1} - \alpha_t) - \beta_{t+1}, \tag{2.80}$$

and integrating only with respect to α_t. Furthermore, since $w_{\lambda,t}(\lambda_t)$ is expected to be near constant over the support of λ_t, we can safely rely upon a "constant weight" approximation whereby the ratio $w_{\lambda,t}(\lambda_t)/\ell_t$ in (2.78) is set equal to 1. Whence $f(\lambda_{t+1}|Y_t)$ can be accurately evaluated by the following bivariate integral:

$$f(\lambda_{t+1}|Y_t) = \int \frac{g_t(\theta_t)}{\chi_t(\theta_t)} k_{\lambda,t}(\lambda_t) f(\alpha_{t+1}|\lambda_t)|_{\theta_t=\theta_t(\alpha_t),\beta_t=\phi^{-1}(\lambda_{t+1},\alpha_t)} d\alpha_t. \tag{2.81}$$

Numerically efficient evaluation of this integral requires the following additional steps:

(i) Combine analytically $k_{\lambda,t}(\lambda_t)$ and $f(\alpha_{t+1}|\lambda_t)$ into a Gaussian kernel in $(\alpha_{t+1}, \lambda_t)$;

(ii) Introduce the transformation from α_t into (ρ_t, θ_t) with Jacobian $\rho_t > 0$;

(iii) Given $(\lambda_{t+1}, \theta_t)$, integrate analytically in $\rho_t > 0$;

(iv) Given λ_{t+1}, use $g_t(\theta_t)$ as a natural sampler and compute the integral using the draws of θ_t obtained in the previous round.

Note that the sequence of operations just described must be repeated for any value of λ_{t+1} for which $f(\lambda_{t+1}|Y_t)$ is to be evaluated for period-$(t+1)$ EIS evaluation of ℓ_{t+1}. However, as illustrated below, the numerical efficiency of the EIS procedures we have just described results in dramatic reductions in the number of MC draws required to reach a preassigned level of numerical accuracy, and thus in significant reductions in overall computing time relative to the particle filter.

Filtered Values

Filtered values for $\{\lambda_t\}_{i=1}^N$ are defined as

$$E(\lambda_t|Y_t) = \int \lambda_t f(\lambda_t|Y_t) d\lambda_t. \tag{2.82}$$

Substituting (2.78) for $f(\lambda_t|Y_t)$ and introducing the transformation from λ_t to s_t produces the following operational expression for the filtered values of λ_t:

$$E(\lambda_t|Y_t) = \frac{1}{\ell_t}\int h(s_t)w_{s,t}(s_t)\,g_t(\delta_t|\theta_t)g_t(\theta_t)ds_t. \tag{2.83}$$

EIS estimates of these filtered values obtain as

$$\widehat{E(\lambda_t|Y_t)} = \frac{\sum_{i=1}^{N} h(\widetilde{s}_t^{0,i})w_{s,t}\left(\widetilde{s}_t^{0,i}\right)}{\sum_{i=1}^{N} w_{s,t}\left(\widetilde{s}_t^{0,i}\right)}. \tag{2.84}$$

Filtered values of s_t are obtained by replacing $h(s_t)$ by s_t in (2.83) and (2.84). Note that, in contrast with the evaluation of $f(\lambda_{t+1}|Y_t)$, we do not implement here the "constant weight" approximation under which $w_{\lambda,t}(\lambda_t)/\ell_t$ is be set to 1 in (2.83) and filtered values simplify into the arithmetic mean of $\left\{h\left(\widetilde{s}_t^{0,i}\right)\right\}_{i=1}^{N}$. As discussed, e.g., by Geweke (1989), the reason for preferring the ratio form in (2.84) is that the use of CRNs in the evaluation of the numerator and denominator typically induces positive correlation between their respective MC estimates, thereby reducing further the MC variance of the ratio.

A non-degenerate version of the problem

The singularity of the transition in (2.61) is a (spurious) consequence of a model specification that assumes measurements at each division point of the grid used for discretization of the random walk for speed. We now consider the case in which a finer grid for discretization is used relative to that used for measurement, while also allowing for measurements made at varying time intervals.

For ease of notation, we focus on two successive measurements separated by D discretization intervals. Equation (2.61) then must be transformed into a transition density for $\lambda_{t+D}|\lambda_t$ by implicit marginalization with respect to the state sequence $\{\lambda_{t+j}\}_{j=1}^{D-1}$. The random walk process for speed is given by

$$\beta_{t+1} = \beta_t + \varepsilon_{t+1}, \qquad \varepsilon_t \sim N\left(0, \sigma^2 I_2\right), \tag{2.85}$$

and position is discretized as

$$\alpha_{t+1} = \alpha_t + \frac{1}{2}(\beta_t + \beta_{t+1}). \tag{2.86}$$

It follows that

$$\beta_{t+D} = \beta_t + u_{t+D}, \tag{2.87}$$

$$\alpha_{t+D} = \alpha_t + D\beta_t + \upsilon_{t+D}, \tag{2.88}$$

with

$$u_{t+D} = \sum_{j=1}^{D} \varepsilon_{t+j}, \qquad \upsilon_{t+D} = \frac{1}{2}\sum_{j=1}^{D} [2(D-j)+1]\,\varepsilon_{t+j}. \tag{2.89}$$

The covariance matrix of $(u_{t+D}, \upsilon_{t+D})$ obtains by application of standard formulae for the sums and sums of squares of natural numbers - see e.g. Gradshteyn and M.Ryzhik (1965). It follows that the transition density from λ_t to λ_{t+D} is given by

$$\lambda_{t+D}|\lambda_t \sim N\left(A_D\lambda_t, \sigma^2 V_D\right), \tag{2.90}$$

with

$$A_D = \begin{pmatrix} I_2 & DI_2 \\ 0 & I_2 \end{pmatrix}, \qquad V_D = D \cdot \begin{pmatrix} \frac{4D^2-1}{12} I_2 & \frac{D}{2} I_2 \\ \frac{D}{2} I_2 & I_2 \end{pmatrix}. \tag{2.91}$$

The case $D = 1$ obviously coincides with the degenerate transition in (2.61). The generalization from (2.61) to (2.90) does not affect EIS evaluation of the likelihood function. However, the evaluation of $f(\lambda_{t+D}|Y_t)$ now requires four-dimensional integration, and is given by

$$f(\lambda_{t+D}|Y_t) = \int \frac{g_t(\theta_t)}{\chi_t(\theta_t)} k_{\lambda,t}(\lambda_t)\, f(\lambda_{t+D}|\lambda_t)\, d\lambda_t. \tag{2.92}$$

The numerical evaluation of equation (2.92) parallels that of equation (2.81) with the additional (analytical) integration with respect to β_t.

Application

We demonstrate our methodology in an application designed essentially along the lines of that constructed by Gordon et al. (1993), and modified by Pitt and Shephard (1999). For the singular and non-singular cases, σ in (2.61) and (2.90) is set to 0.001; and r in (2.63)

is set to $1-(0.005)^2$. The initial latent vector λ_1 is normally distributed with mean vector $(-0.05, 0.2, 0.001, -0.055)$ and diagonal covariance matrix with standard deviations

$(0.05, 0.03, 0.0005, 0.001)$. In the non-singular case, the number D_t of discretization intervals between measurements t and $t+D$ is drawn from a multinomial $\{2, 3, ..., 11\}$ with equal probabilities $p_i = 0.1$, $i = 2, ..., 11$. (Actually, the non-singular case need not be restricted to integer values of D_t, and we have verified that solutions for the non-singular case converge to those for the singular case as D_t tends towards 1.)

We set $T = 10$, and draw two sets of latent vectors $\{\lambda_t^s\}_{t=1}^{10}$, one for the singular case ($s = 1$) and one for the non-singular case ($s = 2$). Both sets are linear transformations of a single set of $N(0,1)$ draws.

As in Pitt and Shephard (1999), we draw $R = 40$ different data sets $\left\{Y_T^{s,i}\right\}_{i=1}^{40}$ based on the latent vectors $\{\lambda_t^s\}_{t=1}^{10}$ for $s = 1, 2$. For each data set, we produce 100 i.i.d. estimates of the filtered means (differing by the seeds initializing the MC draws) using EIS-1K and PF-40K (computing times associated with these procedures are similar with PF-40K requiring 0.439 seconds and EIS-1K 0.492 seconds per function evaluation). As in the SV example, comparing estimates generated by these procedures with "true" filtered means for the latent variables yields LMSE comparisons analogous to those employed by Pitt and Shephard (1999) to demonstrate the gains in precision and efficiency yielded by their extensions of the particle filter. (Details regarding the construction of LMSEs in this case correspond precisely with those described for the SV application.)

Graphs of MC variances and squared biases are presented in Figure 2.3, and LMSEs in Figure 2.4, both for the singular case (similar results were obtained for the non-singular case, and thus are not reported). Note the large reductions in LMSEs yielded by the move from PF to EIS estimates: differences average between 4 and 6 on the log scale. These differences are once again much larger than those reported by Pitt and Shephard (1999): their auxiliary particle filter yielded reductions averaging between 0.5 and 1 relative to the particle filter.

Regarding the source of the large differences in LMSEs we obtain, differences in logged variances are typically of the order of 2 to 2.5 in favor of EIS (corresponding roughly to a

10-fold reduction in variance), except for $t = 1$. Logged bias ratios are virtually all close to zero for EIS filter, while they typically lie between 1 and 4 (and as high as 10 for $t = 1$) for the particle filter. Thus biases remain significant for the particle filter even using 40K draws, and are the dominant component of the large differences in LMSEs generated by the adoption of EIS. This is a manifestation of the "sample impoverishment" problem that results from the very tight distribution of $\lambda_t|Y_t$ relative to that of $\lambda_t|Y_{t-1}$ along the θ_t dimension.

2.6 CONCLUSION

We have proposed an efficient means of facilitating likelihood evaluation and filtering in applications involving non-linear and/or non-Gaussian state space representations: the EIS filter. The filter is adapted using an optimization procedure designed to minimize numerical standard errors associated with targeted integrals. Resulting likelihood approximations are continuous in underlying likelihood parameters, greatly facilitating the implementation of ML estimation procedures. Implementation of the filter is straightforward, and the payoff of adoption can be substantial.

2.7 TABLES AND FIGURES

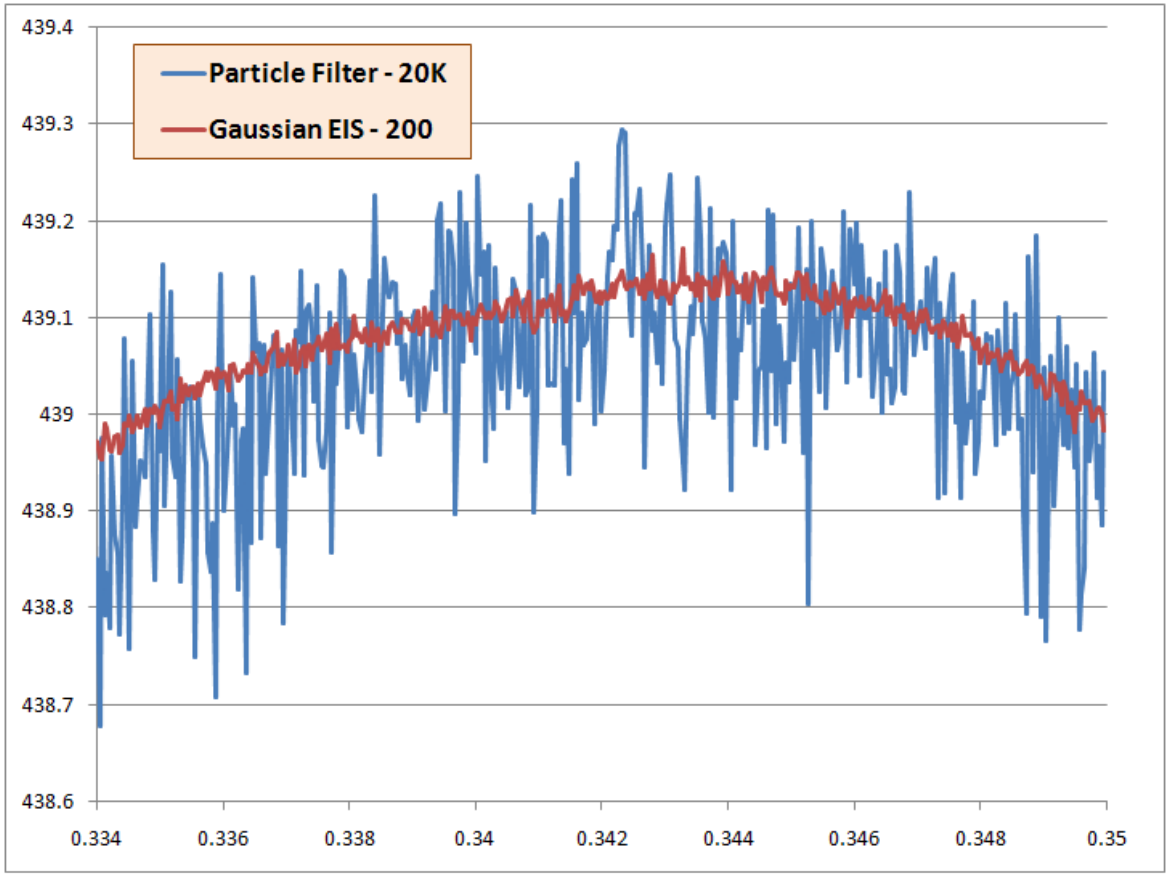


Figure 2.1: Conditional Log Likelihood for α

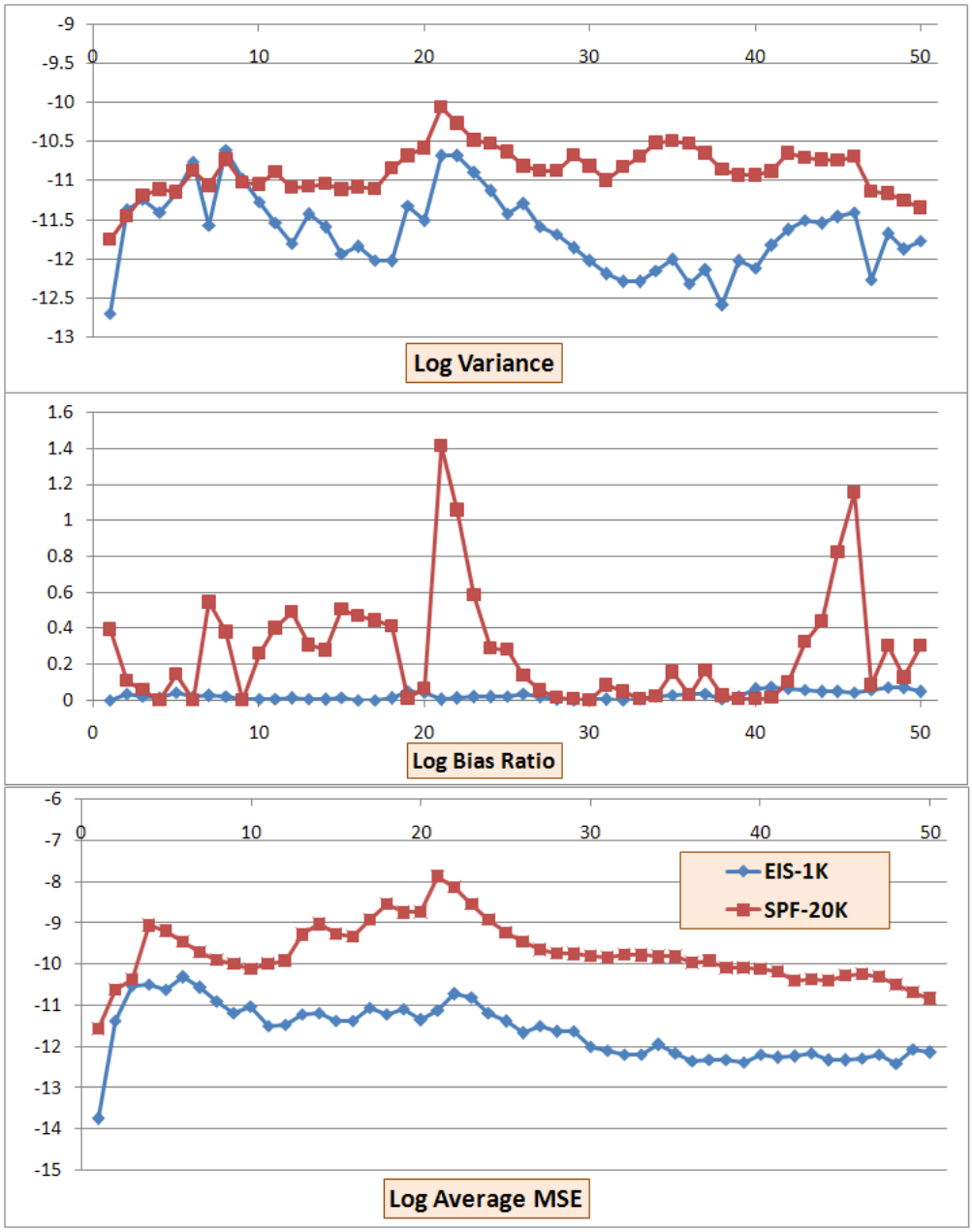


Figure 2.2: MSE decompositions, Stochastic Volatility Model.

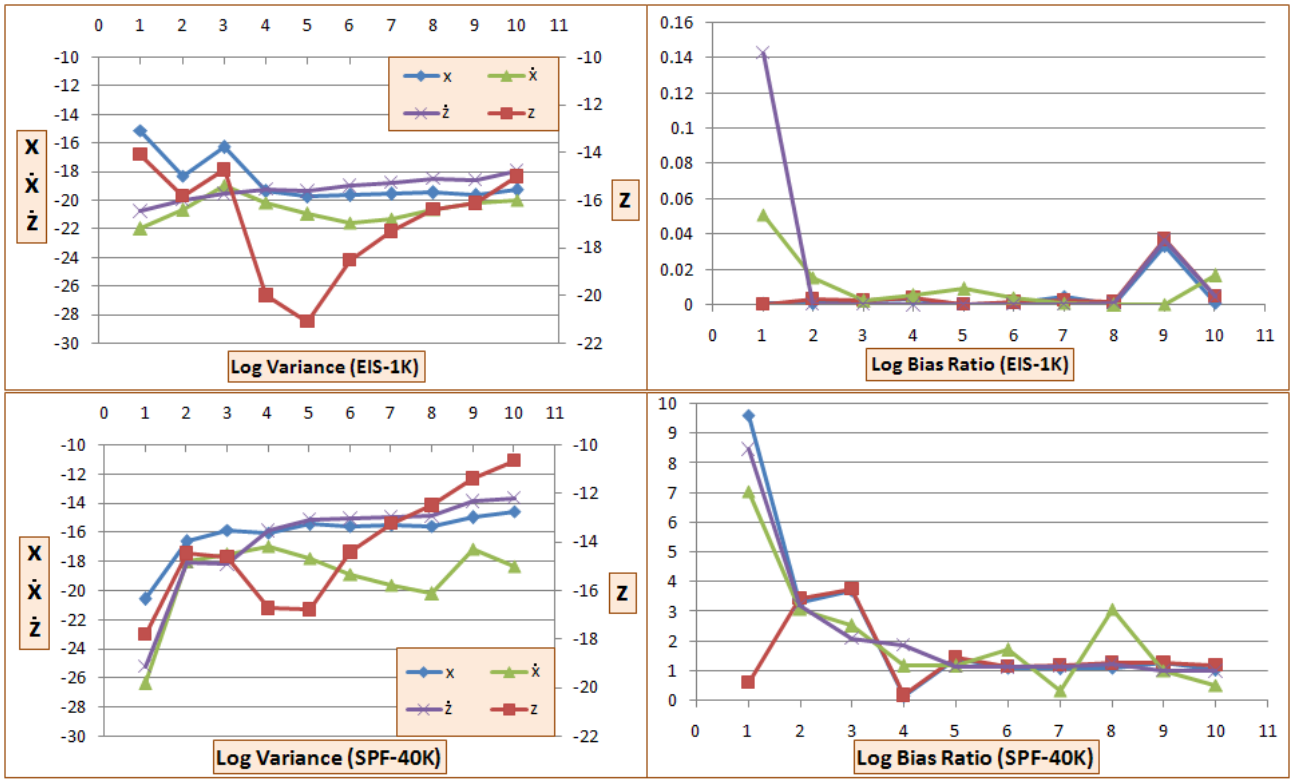


Figure 2.3: Log Variance and Log Bias Ratio, Singular Case.

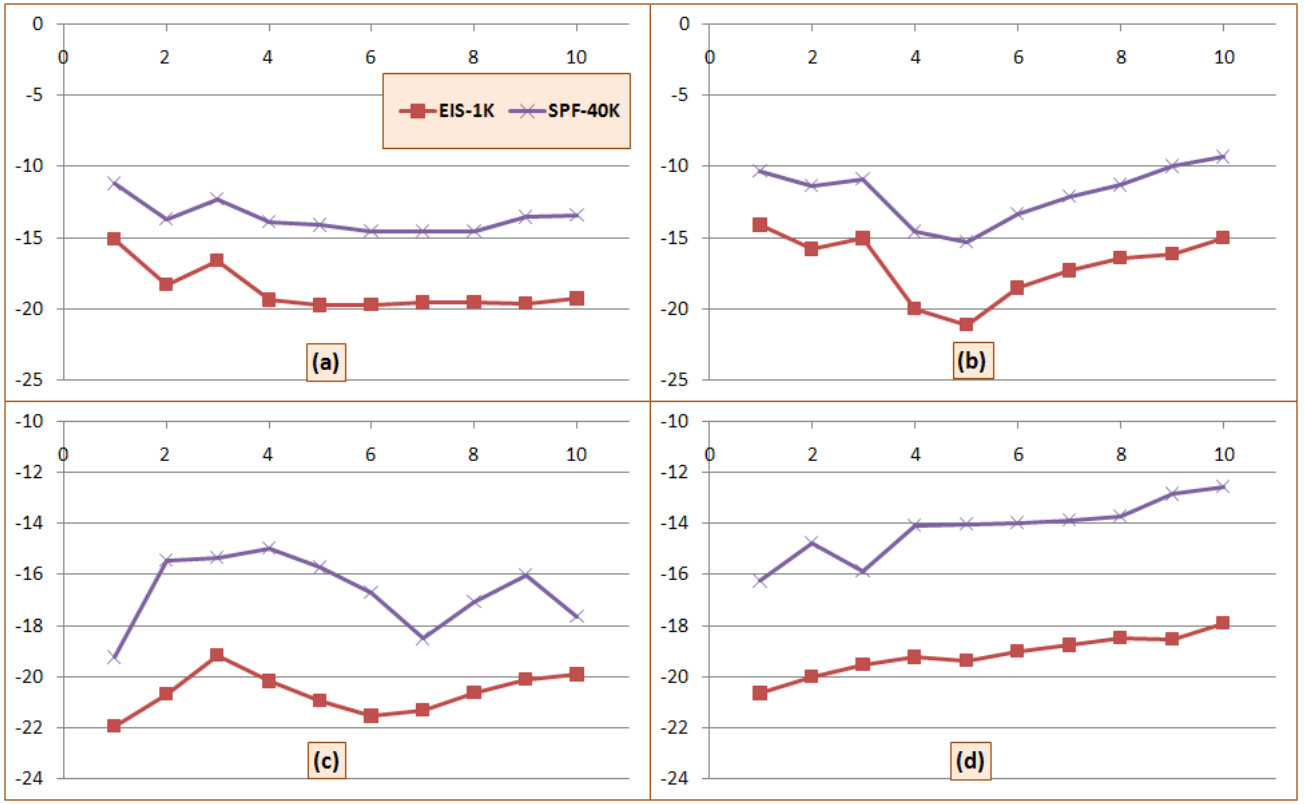


Figure 2.4: Log-Avg.MSE Comparisons, Singular Case. Panel (a) $\rightarrow x$, (b)$\rightarrow z$, (b)$\rightarrow \dot{x}$, (b)$\rightarrow \dot{z}$.

Table 2.1: Table 1. Univariate Model with Frequent Outliers, $u_t \sim t(2)$

STANDARD PARTICLE FILTER												
	$\sigma_\upsilon = \frac{1}{3}$			$\sigma_\upsilon = 1$			$\sigma_\upsilon = 3$			$\sigma_\upsilon = 10$		
N	Mean	Stdv.	Time	Mean	Stdv.	Time	Mean	Stdv.	Time	Mean	Stdv.	Time
20000	-195.846	0.0183	2.065	-213.593	0.0512	2.083	-263.541	0.1089	2.054	-368.542	0.2117	1.995
200000	-195.843	0.0057	22.119	-213.591	0.0159	22.593	-263.534	0.0339	24.269	-368.532	0.0700	22.798
AUXILIARY PARTICLE FILTER												
20000	-195.839	0.0162	2.907	-213.605	0.0461	2.912	-263.531	0.1117	2.930	-368.545	0.2285	2.906
200000	-195.840	0.0055	31.811	-213.601	0.0159	32.343	-263.526	0.0335	32.075	-368.520	0.0697	32.285
Relative Time Efficiency			0.748			0.699			0.775			0.713
ADAPTED PARTICLE FILTER												
20000	-195.839	0.0059	2.710	-213.600	0.0306	2.730	-263.635	0.3902	2.740			
200000	-195.840	0.0017	29.945	-213.604	0.0113	29.935	-263.543	0.1312	29.778			
Relative Time Efficiency			8.182			1.497			0.0543			
PIECEWISE-EIS PARTICLE FILTER (R=100, S=100)												
100	-195.845	0.0409	0.552	-213.595	0.0788	0.5392	-263.530	0.0344	0.549	-368.516	0.0107	0.550
1000	-195.845	0.0139	2.262	-213.602	0.0261	2.180	-263.526	0.0108	2.228	-368.521	0.0049	2.241
Relative Time Efficiency			1.644			3.846			107.32			2001.93
GAUSSIAN-EIS PARTICLE FILTER (R=100, S=100)												
1000	-195.825	0.043	0.717	-213.815	0.145	0.860	-266.016	0.459	1.699	-372.937	0.5117	2.278
1000	-195.819	0.038	2.530	-213.601	0.087	2.287	-264.156	0.459	4.265	-370.858	0.5110	4.491
Relative Time Efficiency			0.199			0.325			0.031			0.095

Note: Means and standard deviations are based on 100 Monte Carlo replications; Relative Time Efficiency is based on N=200,000 for the SPF and N=1,000 for the EIS filter.

Table 2.2: Table 2. Univariate Model with Frequent Outliers, $u_t \sim t(50)$

STANDARD PARTICLE FILTER												
	$\sigma_\upsilon = \frac{1}{3}$			$\sigma_\upsilon = 1$			$\sigma_\upsilon = 3$			$\sigma_\upsilon = 10$		
N	Mean	Stdv.	Time	Mean	Stdv.	Time	Mean	Stdv.	Time	Mean	Stdv.	Time
20000	-141.169	0.0240	2.127	-166.151	0.0687	2.134	-241.043	0.1166	2.053	-356.860	0.2797	2.026
200000	-141.166	0.0077	23.131	-166.149	0.0195	22.915	-241.038	0.0350	22.485	-356.821	0.0750	22.533
AUXILIARY PARTICLE FILTER												
20000	-141.138	0.0199	2.878	-166.153	0.0564	2.806	-241.052	0.1203	2.794	-356.879	0.2464	2.803
200000	-141.139	0.0064	31.423	-166.147	0.0172	31.752	-241.037	0.0350	31.806	-356.811	0.0763	31.234
Relative Time Efficiency			1.067			0.929			0.707			0.697
ADAPTED PARTICLE FILTER												
20000	-141.168	0.0063	2.629	-166.144	0.0425	2.643						
200000	-141.170	0.0020	29.892	-166.145	0.0159	30.015						
Relative Time Efficiency			11.590			1.157						
PIECEWISE-EIS PARTICLE FILTER (R=100, S=100)												
100	-141.167	0.0485	0.5156	-166.147	0.0727	0.568	-241.046	0.0325	0.513	-356.803	0.0173	0.587
1000	-141.164	0.0167	2.0878	-166.128	0.0248	2.324	-241.038	0.0109	2.099	-356.796	0.0061	2.388
Relative Time Efficiency			2.330			6.075			109.97			1401.81
GAUSSIAN-EIS PARTICLE FILTER (R=100, S=100)												
1000	-141.171	0.051	0.833	-166.160	0.087	0.791	-241.088	0.034	0.865	-356.868	0.0215	0.896
1000	-141.175	0.029	2.049	-166.159	0.047	1.996	-241.046	0.020	2.576	-356.812	0.0116	3.308
Relative Time Efficiency			0.782			1.976			25.493			284.197

Note: Means and standard deviations are based on 100 Monte Carlo replications; Relative Time Efficiency is based on N=200,000 for the SPF and N=1,000 for the EIS filter.

Table 2.3: DSGE Model

STANDARD PARTICLE FILTER				
N	Mean	Stdev	Time	Rel. Time Efficiency
20000	438.561	0.206	5.7651	1.000
100000	438.545	0.0774	29.036	1.417
GAUSSIAN-EIS PARTICLE FILTER ($R = 100$)				
200	438.621	0.0278	5.731	55.440
1000	438.633	0.0083	16.414	217.728

Note: Means and standard deviations are based on 100 Monte Carlo replications; Relative Time Efficiency is based on N=20,000 for the SPF.

Table 2.4: MLE Comparisons

.								
STANDARD PARTICLE FILTER (N=20000)								
	True	Stat. Moments				Num. Moments		
T=40	True	Mean	*S.D.*[a]	*S.D.*[b]	RMSE	Mean	*S.D.*[c]	*S.D.*[d]
α	0.33	0.34559	5.888E-03		1.667E-02	0.34826	3.158E-03	5.765E-04
β	0.96	0.91989	1.843E-02		4.414E-02	0.93891	8.490E-03	1.550E-03
ρ	0.8	0.82840	2.899E-02		4.058E-02	0.81292	1.735E-02	3.168E-03
σ	0.05	0.04802	4.363E-03	n/a	4.792E-03	0.05184	1.384E-03	2.527E-04
σ_l	0.014	0.01462	4.118E-03		4.165E-03	0.01588	6.831E-04	1.247E-04
σ_i	0.02	0.01955	2.030E-03		2.080E-03	0.02149	4.275E-04	7.804E-05
T=100								
α	0.33	0.33404	5.701E-03		6.987E-03	0.33922	3.027E-03	5.527E-04
β	0.96	0.91995	1.568E-02		4.301E-02	0.94024	9.035E-03	1.650E-03
ρ	0.8	0.79697	1.961E-02		1.984E-02	0.82860	1.698E-02	3.100E-03
σ	0.05	0.05413	3.348E-03	n/a	5.316E-03	0.05085	1.246E-03	2.275E-04
σ_l	0.014	0.01350	2.958E-03		3.001E-03	0.01416	6.052E-04	1.105E-04
σ_i	0.02	0.01991	1.259E-03		1.262E-03	0.02030	3.544E-04	6.470E-05
T=500								
α	0.33	0.33162	5.440E-03		5.739E-03	0.33399	3.498E-03	6.386E-04
β	0.96	0.95501	1.523E-02		1.611E-02	0.95049	9.930E-03	1.813E-03
ρ	0.8	0.81776	1.599E-02		2.459E-02	0.80170	1.759E-02	3.212E-03
σ	0.05	0.05238	2.579E-03	n/a	3.717E-03	0.05365	1.899E-03	3.468E-04
σ_l	0.014	0.01361	1.717E-03		1.849E-03	0.01400	4.898E-04	8.942E-05
σ_i	0.02	0.01971	5.413E-04		6.368E-04	0.01922	2.604E-04	4.754E-05
GAUSSIAN-EIS PARTICLE FILTER (N=200)								
T=40	True	Mean	*S.D.*[a]	*S.D.*[b]	RMSE	Mean	*S.D.*[c]	*S.D.*[d]
α	0.33	0.34071	2.389E-03	1.974E-03	1.097E-02	0.34868	1.652E-03	3.016E-04
β	0.96	0.93364	1.085E-02	n/a	2.851E-02	0.94006	4.442E-03	8.110E-04
ρ	0.8	0.81669	1.562E-02	1.145E-02	2.286E-02	0.81811	9.791E-03	1.788E-03
σ	0.05	0.04879	3.193E-03	3.009E-03	3.416E-03	0.05425	9.340E-03	1.705E-03
σ_l	0.014	0.01535	2.276E-03	2.107E-03	2.646E-03	0.01594	2.415E-04	4.410E-05
σ_i	0.02	0.01984	1.586E-03	1.544E-03	1.594E-03	0.02155	2.349E-04	4.289E-05
T=100								
α	0.33	0.33380	4.635E-03	3.977E-03	5.996E-03	0.33601	1.577E-03	2.880E-04
β	0.96	0.92038	1.557E-02	n/a	4.337E-02	0.93960	4.441E-03	8.109E-04
ρ	0.8	0.79867	1.764E-02	1.432E-02	1.769E-02	0.82182	9.667E-03	1.765E-03
σ	0.05	0.05083	3.284E-03	3.312E-03	3.388E-03	0.05041	9.492E-03	1.733E-03
σ_l	0.014	0.01407	2.998E-03	2.919E-03	2.999E-03	0.01448	2.471E-04	4.512E-05
σ_i	0.02	0.01990	1.163E-03	1.214E-03	1.167E-03	0.02183	2.144E-04	3.914E-05
T=500								
α	0.33	0.33032	2.235E-03	2.243E-03	2.385E-03	0.33095	1.395E-03	2.547E-04
β	0.96	0.95610	7.478E-03	n/a	8.830E-03	0.95744	7.988E-03	1.458E-03
ρ	0.8	0.81082	6.408E-03	6.199E-03	1.275E-02	0.80102	7.142E-03	1.304E-03
σ	0.05	0.05108	1.064E-03	1.088E-03	1.559E-03	0.05031	4.068E-03	7.428E-04
σ_l	0.014	0.01400	7.206E-04	6.825E-04	7.623E-04	0.01400	2.193E-04	4.003E-05
σ_i	0.02	0.01997	2.295E-04	2.262E-04	2.444E-04	0.01998	1.891E-04	3.453E-05

a. Finite Sample S.D., b. Asymptotic S.D., c. S.D. of a single Draw, d. S.D. of the mean.

3.0 EFFICIENT IMPORTANCE SAMPLING WITH COPULAS

3.1 INTRODUCTION

Dynamic econometric models with latent variables are extensively utilized in economics. Expressed in the state-space form, they are characterized by a transition process describing the evolution of a set of underlying latent (state) variables and a set of measurement equations that relate the observations to the latent variables. For example, state-space models are utilized in macroeconomic research of real business cycle activity (e.g. Dynamic Stochastic General Equilibrium or DSGE models Markov-Switching models for business-cycle turning points etc.)

Likelihood evaluation for parameter estimation and conditional moments of the state variables (filtering/smoothing) requires the integration of the joint density of observable and state variables with respect to the state variables. In the context of linear models with Gaussian errors, these integrals can be computed analytically. With the growing interest in sophisticated nonlinear and non-Gaussian models, development of widely applicable, efficient and robust numerical integration methods has become an important issue in current research. For example, methods like the extended Kalman filter, the second-order filter, Fourier series based methods, Gaussian sum filter etc., approximate non-Gaussian densities with parametric functions or several Gaussian densities. Kitagawa (1987) proposed a non-Gaussian filter and smoother to compute the marginal posterior density of the state variables which approximates the target density with a piecewise approximation.

The development of a Monte-Carlo filter and smoother (called the "Particle Filter", see

Kitagawa 1996 and Gordon, Salmond, and Smith 1993) represents an important advance in the likelihood-based analysis of very general state-space models. Arbitrary non-Gaussian densities in the likelihood-integral are approximated by many "particles" which are considered to constitute a discrete approximation to these distributions. These discrete approximations also serve as importance samplers for the Monte-Carlo (MC) integration over the state-space. As noted by Kitagawa (1998) and Fernandez-Villaverde and Rubio-Ramirez (2007), the particle filter is susceptible to sampling errors (especially in the presence of outliers), causing large errors and non-smoothness of the computed likelihood values. DeJong, Hariharan, Liesenfeld, and Richard (2007) create a new approach to filtering and likelihood evaluation by combining efficient methods for constructing importance sampling densities (called "Efficient Importance Sampling" or EIS, see Richard and Zhang, 2007) with the sequential structure of the particle filter to create a new class of methods called "EIS-Filter". The superior accuracy and smoothness of likelihood values obtained with the EIS-Filter is due to fact that the importance sampling densities closely "mimic" the integrands, minimizing the numerical error of the computed likelihood-integral.

However, in many modeling scenarios, standard parametric densities (e.g. Gaussian) may not serve as good importance samplers. In multivariate problems of moderate dimensions, one or more of the constituent univariate margins may exhibit pathologies like asymmetric tails, multi-modality etc. Thus, it may be desirable to create a multivariate importance sampler by combining (possibly) disparate univariate distributions for the margins into a joint distribution. Copula distributions are perfectly suited for this objective.

A Copula is a function that connects several univariate marginal distributions of individual random variables to form a multivariate distribution. A very readable introduction to copula distributions can be found in Trivedi and Zimmer (2005). Importantly, the choice of marginal distributions need not affect the choice of the copula. Also, the marginal distributions can be arbitrarily dissimilar (possibly belonging to different families of distributions), making the copula a very flexible method for characterizing joint distributions.

Copulas have been extensively utilized in time-series applications in both finance and

economics. A broad review of copula-based models for both multivariate and univariate time-series can be found in Patton (2007). Beyond time-series applications, copulas have also been used in applications involving quantile regressions (Bouyé and Salmon 2002), count data (Heinen and Rengifo 2003), model selection (Smith 2003), a semiparametric study of auctions (Brendstrup and Paarsch 2007), semiparametric density approximation for bounded data (Bouezmarni and Rombouts 2007) etc. Other references include Trivedi and Zimmer (2006), Chib and Winkelmann (2001) and Cameron, Li, Trivedi, and Zimmer (2004).

The flexibility afforded by copula distributions make them ideally suited to serve as importance sampling densities for the MC integration of the likelihood integral in in very general state-space models. As noted by Patton (2007), copula based methods are not suitable for high-dimensional problems. Creating flexible and parsimonious high-dimensional copula distributions is an important area of current research. However, copula densities are a natural choice when maximum flexibility is desired in capturing the marginal behavior of a few (say one to four) univariate marginal random variables. Creating copula-based importance sampling densities also provides the advantage of splitting the density approximation problem into two components: modeling marginal behavior and then (as a second step) modeling the dependence.

The primary focus of this paper is to illustrate the utility of copula-based importance samplers and compare its performance to the particle filter via simulation studies. Building upon the EIS-Filter developed by DeJong et al. (2007), this paper develops computational methods to construct importance sampling densities using copulas. A pilot example involving the integration of a mixture of two bivariate normal random variables serves to illustrate the flexibility of copula-based importance sampling densities in capturing pathologies like asymmetric tails, bimodality etc. in the behavior of the marginal random variables. Also, the copula-based EIS-Filter is utilized to estimate a dynamic stochastic general equilibrium (DSGE) model involving investment-specific technological change (see Fisher, 2003 and Fernandez-Villaverde and Rubio-Ramirez 2004a and 2007) and varying degrees capital utilization (see Chatterjee 2005).

The remaining paper is organized as follows. The next section provides a brief introduction to copulas and methods to estimate the dependence parameters. Section 3.3 provides a brief introduction to state-space models and associated likelihood functions. Efficient Importance Sampling and the use of copulas to construct importance sampling densities are illustrated in Section 3.4. Section 3.5 presents the two examples illustrating the use of copula-based EIS-filter and comparing its performance to the EIS-filter with Gaussian samplers (Gaussian-EIS) and the particle filter.

3.2 A PRIMER ON COPULAS

3.2.1 Basic Properties

Copulas provide a general approach for constructing joint distributions based on marginals from possibly different families. Copulas are important in the study of dependence between random variables since they allow us to separate the effect of dependence from effects of the marginals distributions.

In simple terms, copulas are distribution functions with uniform marginals. The theorem by Sklar (1973) (original article in French was published in 1959) establishes some fundamental properties of copulas as joint distributions. An m-dimensional function $\mathbf{C} : [0,1]^m \rightarrow [0,1]$ is a distribution function if it satisfies the following criteria:

1. $\mathbf{C}(1,...,u_i,1,...,1) = u_i \; \forall i \leq m$ and $u_i \in [0,1]$;
2. For any $i \leq m$, $\mathbf{C}(u_1,...,u_m) = 0$ if $u_i = 0$;
3. $\mathbf{C}$ is m-increasing. (The $\mathbf{C}$-volume of any m-dimensional interval is non-negative).

Let $\mathbf{X} = [X_1,...X_m]'$ be an m-dimensional random variable with joint distribution $\mathbf{H}$ and continuous marginal distributions $F_1,...,F_m$. Sklar's theorem shows that there exists a function $\mathbf{C}(F(X_1;\theta_1),...,F(X_m;\theta_m);\theta_c)$ that links the individual distribution functions to the joint distribution $\mathbf{H}$. Here, θ_i denotes the parameters associated with margin-i and θ_C represents the

parameters associated with the copula alone. The joint distribution is given by

$$\mathbf{H}(x_1,...,x_m;\theta_H) = \mathbf{C}\left(F(X_1;\theta_1),...,F(X_m;\theta_m);\theta_c\right), \tag{3.1}$$

where θ_H represents the parameters of the distribution $\mathbf{H}$. Since $F(x_i;\theta_i) \equiv u_i \sim Uniform(0,1)$ for all i, the copula can also be written as

$$\mathbf{H}(x_1,...,x_m;\theta_H) = \mathbf{C}(u_1,...,u_m;\theta_C), \tag{3.2}$$

illustrating the connection between the univariate random variables X_i and the corresponding margin of the copula u_i.

Assuming that the copula is m times differentiable, the density can be written as

$$h(x_1,...,x_m;\theta_H) = f\left(x_1;\theta_1\right)...f(x_m;\theta_m)c(u_1,...,u_m;\theta_C) \tag{3.3}$$

where

$$c(u_1,...,u_m;\theta_C) = \frac{\partial^m}{\partial u_1...\partial u_m}\mathbf{C}(u_1,...,u_m;\theta_C). \tag{3.4}$$

Hence, the joint density is simply the product of marginal densities and the copula density $c(u_1,...,u_m;\theta_C)$. From the above expression for the copula density it is evident that the joint log-likelihood reduces to being the sum of univariate log-likelihoods of each margin and the copula log-likelihood $\ln c(u_1,...,u_m;\theta_C)$.

3.2.2 Basic Methods for Copula Estimation

From the above description, it is apparent that two sets of parameters require estimation: those associated with the margins, and those associated with the copula. A direct approach is to estimate both sets using the full likelihood. However, copulas allow for this estimation problem to be decomposed into two steps: first the marginals, then the dependence parameter(s) (see Trivedi and Zimmer, 2005 for a more detailed discussion). This two-stage method exploits the fact that the copula dependence parameter(s) are independent of the marginals. Also, the marginals may be estimated using parametric or non-parametric methods (see Bouezmarni and Rombouts 2007 for illustrative examples).

As shown by Joe (1997), it is often easier to work with the log-likelihood of each univariate margin. That is, given a set of N draws from the multivariate distribution $\mathbf{H}$, optimal parameters for the density of the univariate margin-i are obtained by maximizing its log-likelihood

$$\widehat{\theta}_i = \arg\min_{\theta_i} \sum_{j=1}^{N} \log f\left(x_i^j; \theta_i\right).$$

With the optimal parameters (for the margins) obtained one margin at a time, the full likelihood can then be optimized with respected to the dependence parameters only. That is, it is possible to maximize the total likelihood holding the marginal parameters fixed:

$$\widehat{\theta}_C = \arg\min_{\theta_C} = \sum_{i=1}^{m} \sum_{j=1}^{N} \log f\left(x_i^j; \widehat{\theta}_i\right) + \sum_{k=1}^{N} c(u_1, \ldots, u_m; \theta_C).$$

This estimation approach is termed as inference functions for the margins (IFM), following Joe (1997).

3.3 STATE-SPACE REPRESENTATIONS AND LIKELIHOOD EVALUATION

A model expressed in state-space form is described by two sets of equations: the state-transition process describing the evolution of the state variables over time and measurement equations that connect the observed variables to the state variables. Let x_t denote a $m \times 1$ vector of observable variables with the sequence $\{x_j\}_{j=1}^{t}$ being denoted as X_t. Correspondingly, let s_t be a $n \times 1$ vector of unobserved state variables, and denote $\{s_j\}_{j=1}^{t}$ as S_t.

With the conditional expectation of the state variables represented as $E(s_t|s_{t-1}, X_{t-1}) = \gamma(s_{t-1}, X_{t-1})$, the state-transition equations can be written as

$$s_t = \gamma(s_{t-1}, X_{t-1}, \upsilon_t), \tag{3.5}$$

with υ_t being a vector of innovations or structural shocks. Let $E(x_t|s_t, X_{t-1}) = \delta(s_t, X_{t-1})$ denote the expectation of the observed variables conditional on concurrent state variables and the past sequence of observables. The measurement equations can be written as

$$x_t = \delta(s_t, X_{t-1}, v_t) \tag{3.6}$$

where v_t represents the measurement errors. The initial state s_0 is assumed to be distributed according to the known density $f_0(s)$. The parameters in $\gamma(\cdot)$, $\delta(\cdot), u_t, \upsilon_t$ are collectively denoted θ. We assume the required conditional independence assumptions that allow us to exclude the lagged components of X in $g(\cdot)$ and $h(\cdot)$. For reasons of brevity, the parameter vector θ is omitted in the following discussion. Given these conditional independence assumptions, the likelihood function $f(X_T)$ can be decomposed into the product

$$f(X_T) = \prod_{t=1}^{T} f(x_t|X_{t-1}), \tag{3.7}$$

where $f(x_1|X_0) \equiv f(x_1)$. The individual components are obtained by the integration of the joint density of x_t and s_t:

$$f(x_t|X_{t-1}) = \int f(x_t, s_t|X_{t-1})\, ds_t. \tag{3.8}$$

$$\int f(x_t|s_t, X_{t-1})\, f(s_t|X_{t-1})\, ds_t. \tag{3.9}$$

Here, $f(s_t|X_{t-1})$ is computed by marginalization of the joint density of s_t and s_{t-1}:

$$f(s_t|X_{t-1}) = \int f(s_t|s_{t-1},X_{t-1})\, f(s_{t-1}|X_{t-1})\, ds_{t-1}. \tag{3.10}$$

Bayes' theorem implies that the density $f(s_{t-1}|X_{t-1})$ can be written as

$$f(s_{t-1}|X_{t-1}) = \frac{f(x_{t-1},s_{t-1}|X_{t-2})}{f(x_{t-1}|X_{t-2})} = \frac{f(x_{t-1}|s_{t-1},X_{t-2})\, f(s_{t-1}|X_{t-2})}{f(x_{t-1}|X_{t-2})}. \tag{3.11}$$

where the numerator ($f(x_{t-1}|s_{t-1},X_{t-2})\, f(s_{t-1}|X_{t-2})$) represents the integrand in the period-$(t-1)$ equivalent of (3.9) and the denominator $f(x_{t-1}|X_{t-2})$, the period-$(t-1)$ likelihood. Given the need for integration in (3.8) and (3.10), the next section provides a brief overview of efficient importance sampling (EIS).

3.4 EIS-FILTER WITH COPULAS

Given the need for integrating a non-negative function of the form

$$\aleph = \int \varphi(x)dx, \tag{3.12}$$

EIS (Richard and Zhang, 2007) is a method for constructing importance sampling densities that are both continuous and global approximations to the integrand. The EIS method begins with the choice of a parametric family of importance sampling densities:

$$\aleph = \int \frac{\varphi(x)}{g(x;a)} g(x;a)dx. \tag{3.13}$$

The parameters that characterize $g(x;a)$ are refered to as the "auxiliary parameters" and optimality of EIS method is achieved by choosing optimal values for the auxiliary parameters obtained as a solution to the minimization problem

$$(\widehat{a},\widehat{c}) = \arg\min_{a,c} \int [\ln\varphi(x) - c_o - \ln k(x;a)]^2\, g(x;a)dx, \tag{3.14}$$

where

$$g(x;a) = \frac{k(x;a)}{\chi(a)},\ \chi(a) = \int k(x;a)dx, \tag{3.15}$$

and c_o is an intercept meant to calibrate $\ln(\varphi/k)$. The objective of the above minimization problem is to minimize the variance of the ratio $\frac{\varphi(x)}{g(x;a)}$ over the range of integration. Since the auxiliary parameters also appear in the sampling density in (3.14), the minimization problem is implemented as a fixed-point problem in a and c_o. That is, given current estimates a^l and c_o^l, updated estimates are obtained as solution to the least-squares problem.

$$(a^{l+1}, c_o^{l+1}) = \arg\min_{a,c} \sum_{i=1}^{R} \left[\ln \varphi(x_i) - c_o - \ln k(x_i;a)\right]^2. \tag{3.16}$$

The draws $\{x_i\}_{i=1}^R$ are obtained using the current sampler $g(x;a^l)$ and this procedure is repeated until convergence in a and c_0. Appendix B provides an extended description of EIS with a Gaussian importance sampler.

Combining (3.16) with (3.3), we can write the least-squares problem as

$$\begin{aligned}(\theta_1^{l+1}, \ldots, \theta_m^{l+1}, \theta_C^{l+1}, c_o^{l+1}) \quad = \quad & \arg\min_{a,c} \sum_{i=1}^{R} [\ln \varphi(x_i) - c_o^l - \ln c(u_1, \ldots, u_m; \theta_C^l) \\ & - \ln f(x_1;\theta_1) \ldots - \ln f(x_m;\theta_m)]^2. \end{aligned} \tag{3.17}$$

where θ_i denotes the parameters associated with margin-i.

The flexibility of copula specifications can be successfully exploited by splitting the above problem into its components: the estimation of auxiliary parameters associated with each marginal density, and the estimation of parameters associated with the copula density θ_C. It is important to note that this decomposition implies some loss of efficiency. Separate application of the above refinement procedure to individual sub-problems may not obtain the optimal properties of the combined approach. However, this potential loss of efficiency is offset by the ability to combine diverse marginal distributions into a multivariate importance sampling density. Also, the piecewise-continuous density approximation introduced in DeJong et al. (2007) can be successfully exploited to create accurate univariate samplers for the margins, while the copula accounts for the dependence.

An issue that is immediately apparent from the above description is the choice of the copula density. It is important to note that this choice is inherently problem-specific. For example, a simple yet flexible copula is the meta-Gaussian distribution (see Kelly and Krzysztofowicz,

1997; Fang, Fang, and Kotz, 2002). The meta-Gaussian density is constructed by combining the normal quantile transformation of the marginal random variables into a multivariate Gaussian density. This density can represent the full range of linear dependence between the co-variates while permitting arbitrary marginal distributions.

Let $\Phi_{\Re}$ represent the CDF of a multivariate Gaussian distribution with zero mean and correlation matrix $\Re$, and Φ represent the univariate Gaussian CDF. The meta-Gaussian distribution is given by

$$H(x_1,...x_m;\Re,\theta_1,...,\theta_m) = \Phi_{\Re}\left(\Phi^{-1}(u_1),...,\Phi^{-1}(u_m);\Re\right) \tag{3.18}$$

where $F(x_i;\theta_i)$ represent the univariate margin-i and $u_i = F(x_i;\theta_i)$. Let $\widetilde{x}_i$ represent the transformed versions of x_i, that is

$$\widetilde{x}_i = \Phi^{-1}(F(x_i;\theta_i)). \tag{3.19}$$

The CDF and PDF of a 2-dimensional meta-Gaussian distribution are

$$H(x_1,x_2;\rho,\theta_1,\theta_2) = \Phi_\rho(\widetilde{x}_1,\widetilde{x}_2;\rho), \tag{3.20}$$

$$h(x_1,x_2;\rho,\theta_1,\theta_2) = \frac{f(x_1;\theta_1)f(x_2;\theta_2)}{\sqrt{1-\rho^2}}\exp\left\{-\frac{\left[\widetilde{x}_1^2 - 2\rho\widetilde{x}_1\widetilde{x}_2 + \widetilde{x}_2^2\right]}{2(1-\rho^2)}\right\}, \tag{3.21}$$

and ρ is the correlation between $\widetilde{x}_1$ and $\widetilde{x}_2$. Thus, (3.20) is a Gaussian CDF describing the transformed variables $\widetilde{x}_1$ and $\widetilde{x}_2$. As emphasized by Kelly and Krzysztofowicz (1997), the meta-Gaussian density imposes no restriction on the marginal distributions and provides a convenient analytical expression for the joint density. Given correlation ρ and the marginal distributions, drawing from the two-dimensional meta-Gaussian density reduces to drawing from a bivariate Gaussian density with correlation ρ to obtain draws of the transformed variates $\widetilde{x}_1$ and $\widetilde{x}_2$; draws from the marginals can be obtained by reversing the transformation in (3.19). A meta-t distribution can be defined analogously using the t distribution; see Demarta and McNeil (2005) for a description of the t-copula and the meta-t distribution.

The following section illustrates the flexibility afforded by EIS with copula densities by presenting a pilot application involving the integration of a mixture of two bivariate normal densities. The accuracy of likelihood evaluation and filtering are evaluated via the likelihood

analysis of a DSGE model. The main objective of this exercise being, the comparison of the smoothness and numerical accuracy (of both likelihood evaluation and filtering) of the copula-based EIS-filter to the particle filter.

3.5 EXAMPLES

3.5.1 Mixture of Bivariate Normals

Asymmetric tails, bimodality and other non-Gaussian behavior can be represented by a mixture of two bivariate normal densities. Let μ_1 and μ_2 represent the mean vectors two bivariate normal distributions with the identity matrix (I_2) as their covariance matrices. Let $p \in (0,1)$ represent the mixture probability. The mixture normal density of x $(= [x_1, x_2]')$ is given by

$$\begin{aligned} f(x;\mu_1,\mu_2) &= p\left[\frac{1}{2\pi}\exp\left(-\frac{1}{2}(x-\mu_1)'(x-\mu_1)\right)\right] \\ &+ (1-p)\left[\frac{1}{2\pi}\exp\left(-\frac{1}{2}(x-\mu_2)'(x-\mu_2)\right)\right]. \end{aligned} \tag{3.22}$$

Integration of such densities using the copula-based importance samplers serves as an ideal showcase for the flexibility afforded by copula densities. As shown in Figures 3.1 and 3.3, Varying the means μ_1 and μ_2 can induce widely different marginal behavior. Setting $\mu_1 = [-1,0]'$, $\mu_2 = [0,2]'$ and $p = 0.3$, we obtain a density with some asymmetric tail behavior as depicted in Figure 3.1. The second specification (illustrated in Figure 3.3) with $\mu_1 = [2,-2]'$, $\mu_2 = [-2,3]'$ and $p = 0.6$ exhibits pronounced bimodality.

The mixture density in both the specifications was integrated using the Gaussian-EIS procedure (see Appendix B for an algorithmic description of Gaussian-EIS) and using a bivariate meta-Gaussian importance sampler. The first specification represents a relatively mild departure from Gaussianity, while the second specification represents a pronounced deviation from Gaussian behavior. Equation (3.22) being a proper density, it integrates to 1 in both cases. The piecewise-continuous density approximation introduced in DeJong et al. (2007) was utilized to approximate the margins of the copula importance sampler. If the margins were also set

to be Gaussian densities, the margins in equation (3.17) would be univariate Gaussian kernels in addition to the copula density $c(\cdot)$ being a bivariate Gaussian kernel. Optimal auxiliary parameters can be either be computed jointly or individually. The auxiliary regression would continue to be of the sort illustrated in Appendix B.

The piecewise density approximation is "optimized" by using an initial approximation to obtain an equal probability partition of the space of the state variable. Thus, with the margins specified to be piecewise continuous densities, auxiliary least squares regressions are solely utilized for the copula parameters (in the current example, ρ).

3.5.1.1 The meta-Gaussian Sampler: Piecewise Margins

We begin by initializing the correlation matrix of the bivariate meta-Gaussian density to be an identity matrix ($\rho = 0$). The margins were set to being uniform densities on sufficiently broad intervals in the space of x_1 and x_2. N draws of $\widetilde{x}$ $(= [\widetilde{x}_1, \widetilde{x}_2]')$ were obtained from a bivariate standard normal density. The draws of $\widetilde{x}$ were transformed into draws of x $(= [x_1, x_2]')$ by the normal quantile transform and the uniform marginals (as shown in equation 3.19).

Defining a uniformly spaced grid (with R intervals) on each margin, the univariate marginal density at each node in the grid was approximated by univariate kernel density approximations using the N draws of x. The computed integrand (normalized) served as weights in the kernel density approximation. The KDEs were obtained using a Gaussian kernel. Both Silverman's rule-of-thumb bandwidth and the Sheather-Jones plug-in bandwidth selectors were used (see Sheather 2004 for a practical description of density estimation using kernels). Practical experience suggests that mild oversmoothing is preferable to undersmoothed kernel density estimates. The rule-of-thumb bandwidth produced density estimates marginally smoother than those using the Sheather-Jones plug-in bandwidth. Also, it was found beneficial to utilize fairly large number of draws for the kernel density approximation (N was set to 1000 in the current example). The value of R was set to 100.

Piecewise-linear approximants to the (log) of the computed density were constructed over the uniformly spaced grid and converted to a density following the procedures outlined in DeJong et al. (2007). This piecewise density was "refined" by inverting a uniformly spaced

grid on the unit interval to obtain an equal-probability division of the domain of integration in the univariate margins. The piecewise-continuous density thus constructed serve as the marginal density in the meta-Gaussian approximation[1] and were not revised any further.

3.5.1.2 The meta-Gaussian Sampler: Correlation Matrix

Having obtained the piecewise density approximations for the margins, the optimal correlation parameter ρ was obtained using the iterated procedure outlined in Section 3.4. Initializing ρ to be zero, 100 (size of auxiliary regression, denoted R in equation 3.16) draws of $\widetilde{x}$ were obtained from a bivariate normal density with zero mean and correlation ρ. Draws of $\widetilde{x}$ were converted to the marginal random variables x using the piecewise-continuous densities constructed above and inverse of the transformation in (3.19).

The coefficient of $\widetilde{x}_1\widetilde{x}_2$ from the linear regression of $logf(x;\mu_1,\mu_2)$ on a constant and the cross-product $\widetilde{x}_1\widetilde{x}_2$ yields an updated value for the correlation parameter ρ. Note that the updated value of ρ is obtained as the solution of a quadratic equation with two real roots, with one of them being between $(-1,1)$ and the other being its reciprocal. The computational procedure is described in Appendix C.

3.5.1.3 Results

Table 3.1 presents the mean, standard deviation of the computed integral, and average computational times obtained using 100 MC replications. In the first scenario (depicted in Figure 3.1), with some asymmetric behavior in the tails, Gaussian-EIS procedure with just 100 MC draws produces accurate results and is much faster than the meta-Gaussian EIS sampler. With the meta-Gaussian EIS Sampler, practical experience suggests the use of relatively large number of draws for the construction of the margins using kernels. The process of creating margins using kernels consumes most of the computational time associated with the meta-Gaussian EIS procedure. The iterative process for obtaining ρ converges within 5 to 10 iterations and requires much less computational time relative to the process for the margins. However, as

[1] Codes for the Sheather-Jones plug-in bandwidth computation can be found at J. S. Marron's web-site at http://www.unc.edu/∼ marron/marron.html

Figure 3.1 indicates, the piecewise-approximations for the margins do reflect asymmetric tails, capturing the non-Gaussian marginal behavior successfully. Figure 3.2 shows that the log-EIS-weights (log of the ratio of the integrand to the sampling density) are clustered around zero with both methods.

The second scenario (depicted in Figure 3.3) clearly illustrates the advantages of the meta-Gaussian EIS sampler. Piecewise approximations for the margins clearly capture the bimodality in the margins. Also, the histogram of the log-EIS-weights in Figure 3.4 shows that the bulk of draws are centered around zero. Since kernel density approximation typically implies some smoothing, the sampler weights the low-probability region between the modes a little higher than the target density; resulting in a small fraction of the draws with lower weights. However, the impact of these low weights is minimal. Gaussian-EIS repeatedly captures one of the two components in the bivariate mixture. Hence, the values of the integral computed using Gaussian-EIS are close to the mixing probabilities used and nowhere near one. This is also reflected in the lower panel of Figure 3.4 where the log-EIS-weights are clustered around 0.5.

A comparison of the computational times in Table 3.1 shows that the meta-Gaussian copula sampler is slower than Gaussian-EIS. Most of the additional time is used for the construction of the margins and reflects the cost of having greater flexibility in accurately capturing highly non-Gaussian behavior of the margins. These results illustrate the advantages of using copulas in constructing importance samplers when flexibility in capturing complicated marginal behavior is critical. The following section illustrates likelihood evaluation, parameter estimation and filtering using a copula-based EIS and contrasts its performance to that of the particle filter.

3.5.2 A DSGE Model

Recent papers in macroeconomic literature highlight the role of investment-specific technological change as an important source of business cycle fluctuations (see Fisher 2003 and, Greenwood, Hercowitz, and Krusell 2000). Two important observations are often cited to

highlight the role of investment-specific technological change: the relative price of capital in terms of consumption goods has decreased nearly every year since the 1950s; and the relative price of capital decreases faster during expansions than during recessions (Fernandez-Villaverde and Rubio-Ramirez, 2004a).

Also of concern is the behavior of capital stock utilization and its implication on growth and convergence of economies (Chatterjee, 2005). While capital accumulation is treated as important for economic growth, there exist few studies that study the implications of the "intensity of use" of capital in a dynamic framework. As mentioned by Chatterjee (2005), decisions regarding intensity of capital utilization affect decisions regarding current output and via its relation to the rate of depreciation, the future capital shock. Combining the approach of Fisher (2003) and Chatterjee (2005), a DSGE model that allows for both investment-specific shocks and capital utilization decision (which affects the depreciation rate) is presented below. The aim is to estimate the parameters of the model by maximizing the associated likelihood function. Also of interest are the optimal policy functions for both consumption and the capital utilization factor, and the interaction of this utilization factor with both kinds of (neutral and technology specific) shocks.

Let C_t, I_t, Y_t, K_t, A_t, V_t represent consumption, investment, output, capital stock, technology-neutral shock, and an investment-specific shock respectively. The model is that of a representative agent who seeks to maximize the expected value of lifetime utility by choosing consumption C_t and the degree of capital utilization u_t

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \log(C_t) \tag{3.23}$$

subject to the resource constraint

$$C_t + I_t = A_t (u_t K_t)^{\alpha}, \quad \alpha \in (0,1). \tag{3.24}$$

The law-of-motion for capital is given by

$$K_{t+1} = (1 - \delta(u_t)) K_t + V_t I_t, \quad K_0 \text{ is given}, \ \delta(u) \in (0,1), \tag{3.25}$$

and the evolution of shocks over time is given by

$$A_t = e^{(\gamma+\varepsilon_a)}A_{t-1}, \quad \gamma \geq 0 \tag{3.26}$$

$$V_t = e^{(\nu+\varepsilon_v)}V_{t-1}, \quad \nu \geq 0 \tag{3.27}$$

where

$$\begin{aligned} \varepsilon_a &\sim N(0,\sigma_a^2) \\ \varepsilon_v &\sim N\left(0,\sigma_v^2\right). \end{aligned}$$

As mentioned by Fernandez-Villaverde and Rubio-Ramirez (2004a), this specification for the shocks allows for the possibility of changes in the long run relative price of capital. The rate of depreciation of the capital stock, $\delta(u_t)$ is dependent on the capital utilization parameter. This dependence is given by

$$\delta(u_t) = du^{\phi}, \phi > 1, \;\; d > 0, \;\; 0 \leq \delta(u) \leq 1, \tag{3.28}$$

where

$$\delta'(u) > 0, \;\; \delta''(u) > 0.$$

Following Fernandez-Villaverde and Rubio-Ramirez (2004a) and Fisher (2003), the model can be transformed into a stationary problem by normalizing with variables that are fully known prior to the realization of the current shocks. Let

$$Z_t = \left(A_{t-1}V_{t-1}^{\alpha}\right)^{\frac{1}{1-\alpha}}, \tag{3.29}$$

and rescale variables by $\breve{Y}_t = \frac{Y_t}{Z_t}$, $\breve{C}_t = \frac{Y_t}{Z_t}$, $\breve{I}_t = \frac{I_t}{Z_t}$, $\breve{A}_t = \frac{A_t}{A_{t-1}}$, $\breve{V}_t = \frac{V_t}{V_{t-1}}$, $\breve{Z}_t = \frac{Z_t}{Z_{t-1}}$ and, $\breve{K}_t = \frac{K_t}{Z_t V_{t-1}}$. Therefore,

$$\begin{aligned} \breve{A}_t &= \exp\left(\gamma+\varepsilon_a\right), \\ \breve{V}_t &= \exp\left(\nu+\varepsilon_\nu\right), \\ \breve{Z}_t &= \exp\left(\frac{\gamma+\alpha\nu+\varepsilon_a+\alpha\varepsilon_\nu}{1-\alpha}\right) \text{ and,} \\ Z_t V_{t-1} &= \left(A_{t-1}V_{t-1}\right)^{\frac{1}{1-\alpha}}. \end{aligned}$$

With the above transformations at hand, the first-order conditions for the transformed problem can be written as follows. The Euler equation is given by

$$\frac{\breve{Z}_{t+1}}{\breve{C}_t} = \beta E_t \frac{1}{\breve{C}_{t+1}} \left[\alpha \breve{A}_{t+1} u_{t+1}^{\alpha} \left(\breve{K}_{t+1} \right)^{\alpha-1} + \frac{1-\delta(u_{t+1})}{\breve{V}_{t+1}} \right]. \tag{3.30}$$

The remaining conditions include

$$\breve{Z}_{t+1} \breve{K}_{t+1} = (1-\delta(u_t)) \frac{\breve{K}_t}{\breve{V}_t} + \breve{X}_t, \tag{3.31}$$

$$d\phi u_t^{\phi-1} = \alpha \breve{A}_t \left(u_t \breve{K}_t \right)^{\alpha-1}, \tag{3.32}$$

and

$$\breve{C}_t + \breve{I}_t = \breve{A}_t \left(u_t \breve{K}_t \right)^{\alpha}. \tag{3.33}$$

The existence of an optimal choice for the capital utilization rate in the above framework is shown by Chatterjee (2005). The optimality condition for u_t in (3.32) equates the marginal benefit of utilizing capital to the marginal depreciation cost. Steady state values of the transformed capital stock, investment and capital utilization rate are computed using the above first-order conditions. The steady state values of $\breve{A}_t$, $\breve{V}_t$, and $\breve{Z}_t$ are given by $\exp\left(\gamma + \frac{\sigma_a^2}{2}\right)$, $\exp\left(\nu + \frac{\sigma_v^2}{2}\right)$ and $\exp\left(\frac{\gamma+\alpha\nu}{1-\alpha} + \frac{\sigma_a^2+\alpha^2\sigma_v^2}{2(1-\alpha)^2}\right)$ respectively. Combining the Euler equation in (3.30) and the first-order condition for capital utilization factor yields the steady state value for capital:

$$\overline{\breve{K}} = \left(\frac{\alpha \overline{\breve{A}}}{d\phi} \right)^{\frac{\alpha}{\phi(1-\alpha)}} \left[\frac{\alpha \overline{\breve{A}} \beta \left(\overline{\breve{V}} \phi - 1 \right)}{\phi \left(\overline{\breve{V}} \overline{\breve{Z}} - \beta \right)} \right]^{\frac{\phi-\alpha}{\phi(1-\alpha)}}. \tag{3.34}$$

The first-order conditions for the capital utilization factor, law of motion for capital, the resource constraint and the production function provide the steady state values for other factors:

$$\overline{u} = \left(\frac{\alpha \overline{\breve{A}}}{d\phi} \right)^{\frac{1}{\phi-\alpha}} \overline{\breve{K}}^{\frac{\alpha-1}{\phi-\alpha}} \tag{3.35}$$

$$\overline{\breve{I}} = \overline{\breve{K}} \left(\frac{\overline{\breve{V}} \overline{\breve{Z}} + d\overline{u}^{\phi} - 1}{\overline{\breve{V}}} \right). \tag{3.36}$$

3.5.2.1 Likelihood Evaluation and Filtering using the Copula EIS-Filter

In the remainder of the study, we work with the transformed (stationary version) of the DSGE model described above. The variables $\breve{A}_t$, $\breve{V}_t$ and $\breve{K}_t$ are treated as unobserved. Output and investment are treated as observables. Let $x_t = [\breve{y}_t, \breve{i}_t]'$ denote the observable variables. The measurement equations of the state-space model are given by

$$\breve{y}_t = \breve{A}_t \left(u_t \breve{K}_t\right)^{\alpha} + \eta_y \tag{3.37}$$

and

$$\breve{i}_t = \breve{X}_t + \eta_i. \tag{3.38}$$

Here, $\eta_y \sim N(0, \sigma_y^2)$ and $\eta_i \sim N(0, \sigma_i^2)$ are the measurement errors. Let $s_t = [\breve{A}_t, \breve{V}_t, \breve{K}_t]'$ denote the state variables.The state-transition equations are given by

$$\breve{A}_t = e^{(\gamma + \varepsilon_a)} \tag{3.39}$$

$$\breve{V}_t = e^{(v + \varepsilon_v)} \tag{3.40}$$

$$\breve{Z}_t \breve{K}_t = (1 - \delta(u_{t-1})) \frac{\breve{K}_{t-1}}{\breve{V}_{t-1}} + \breve{X}_{t-1}. \tag{3.41}$$

The discount factor β, and other model parameters α, ϕ, d, γ, ν, σ_a, σ_v, σ_y, σ_i are collectively referred to as θ. The likelihood function is given by

$$f(X_T; \theta) = \prod_{t=1}^{T} f(x_t | X_{t-1}; \theta), \tag{3.42}$$

where $f(x_1 | X_0; \theta) \equiv f(x_1; \theta)$. The individual components are obtained by the integration of the joint density of x_t and s_t:

$$f(x_t | X_{t-1}; \theta) = \int f(x_t | s_t, X_{t-1}; \theta) f(s_t | X_{t-1}; \theta) \, ds_t. \tag{3.43}$$

Here, the prediction density $f(s_t | X_{t-1}; \theta)$ is computed by marginalization of the joint density of s_t and s_{t-1}:

$$f(s_t | X_{t-1}; \theta) = \int f(s_t | s_{t-1}, X_{t-1}; \theta) f(s_{t-1} | X_{t-1}; \theta) \, ds_{t-1}. \tag{3.44}$$

In order to evaluate the smoothness and numerical accuracy of the copula-based EIS-filter, a simulation study comparing it to the standard particle filter was conducted. The results from these comparisons are presented below[2]. The parameters are estimated via maximum likelihood using the EIS-Filter. Also, the accuracy of the EIS-Filter is compared to that of the standard particle filter in a filtering exercise.

At any period t, computing the likelihood in (3.43) requires the evaluation of the optimal choice for consumption and the optimal value for capital utilization parameter. The optimal consumption function was approximated using Chebyshev polynomial approximations defined over the space of the state variables ($\breve{A}_t$, $\breve{V}_t$ and $\breve{K}_t$) (see Judd, 1998 or Miranda and Fackler, 2002 for details).

3.5.2.2 Evaluation of the Prediction Density

Note that the period-t prediction density (3.44) requires the evaluation of an integral. DeJong et al. (2007) present different methods for computing the prediction density, the simplest being a weighted-sum of the transition density $f(s_t|s_{t-1})$ over draws of the s_{t-1} from the previous period's importance sampling density.

Given values of $\breve{K}_{t-1}$,$\breve{A}_{t-1}$, $\breve{V}_{t-1}$, (3.41) represents a Dirac transition. It is possible to partition the state-space into two parts (denoted p_t and q_t): with p_t representing a set of variables with non-degenerate transitions and $q_t|p_t, s_{t-1}, X_{t-1}$ representing the conditionally non-stochastic transition. In our model specification, the partition comprised of $p_t = [\breve{A}_t, \breve{V}_t]'$ and $q_t = \breve{K}_t$. It is important to note that this partition is not unique. It is possible to partition the state-space into stochastic and non-stochastic components in multiple ways. We can express q_t as

$$q_t = \xi(p_t, p_{t-1}, q_{t-1}). \tag{3.45}$$

More specifically,

$$\breve{K}_t = \xi\left(\breve{A}_t, \breve{V}_t, \breve{A}_{t-1}, \breve{V}_{t-1}, \breve{K}_{t-1}\right).$$

[2] A study treating the capital utilization parameter as an unobserved variable and including time-varying volatilities in the shock processes is under progress.

The policy function for optimal consumption can be used for computing the non-stochastic transition in capital stock and its derivatives. Following DeJong et al. (2007), we interpret (3.45) as the limit of a uniform density for $q_t|p_t, s_{t-1}$, on the interval $[\xi(p_t, s_{t-1}) - \varepsilon, \xi(p_t, s_{t-1}) + \varepsilon]$ with $\varepsilon \to 0$. Under the assumption that $\xi(p_t, s_{t-1})$ is differentiable and strictly monotone in q_{t-1}, we can express the inverse transformation as

$$q_{t-1} = \varsigma(p_t, q_t, p_{t-1}) \tag{3.46}$$

with the associated Jacobian

$$J(p_t, q_t, p_{t-1}) = \frac{\partial}{\partial q_t}\varsigma(p_t, q_t, p_{t-1}). \tag{3.47}$$

Using (3.41), the Jacobian can be written as

$$J(\breve{K}_{t-1}) = \breve{V}_{t-1}\breve{Z}_t \left[1 - \breve{K}_{t-1}\frac{\partial\delta(u_{t-1})}{\partial u_{t-1}}\frac{\partial u_{t-1}}{\partial \breve{K}_{t-1}} - \delta(u_{t-1}) + \frac{\partial \breve{X}_{t-1}}{\partial \breve{K}_{t-1}}\breve{V}_{t-1}\right]^{-1}. \tag{3.48}$$

Exploiting the fact that $\breve{A}_t$ and $\breve{V}_t$ are independent of past $\breve{A}_{t-1}$ and $\breve{V}_{t-1}$, the limit of the integral in (3.45) with $\varepsilon \to 0$, gives us the following integral for the state-transition:

$$f(s_t|X_{t-1};\theta) = f(\breve{A}_t;\theta)f(\breve{V}_t;\theta)\int J\left(\breve{K}_{t-1}\right) f(s_{t-1}|X_{t-1};\theta)\,|_{\breve{K}_{t-1}=\varsigma\left(\breve{K}_t, p_{t-1}\right)} dp_{t-1}. \tag{3.49}$$

For any s_t, evaluation of (3.49) requires the evaluation of $f(s_{t-1}|X_{t-1})$ along the zero-measure subspace $\breve{K}_{t-1} = \varsigma\left(\breve{K}_t, p_{t-1}\right)$. Under the assumption that the EIS-weights are constant, we can replace $f(s_{t-1}|X_{t-1};\theta)$ with the previous period's optimal importance sampler $g(s_{t-1}|\widehat{a}_{t-1})$ to give

$$\widehat{f}(s_t|X_{t-1};\theta) = f(\breve{A}_t;\theta)f(\breve{V}_t;\theta)\int J\left(\breve{K}_{t-1}\right) g(s_{t-1}|\widehat{a}_{t-1})\,|_{\breve{K}_{t-1}=\varsigma\left(\breve{K}_t, p_{t-1}\right)} dp_{t-1}. \tag{3.50}$$

For a given $\breve{K}_t$, and a pair $(\breve{A}_{t-1}, \breve{V}_{t-1})$, the value of $\breve{K}_{t-1}$ can be obtained from (3.41). Since $\delta(u_{t-1})$ and $\breve{X}_{t-1}$ are also dependent on $\breve{K}_{t-1}$, computation of $\breve{K}_{t-1}$ is accomplished using the Newton-Raphson method by solving (3.41). The Jacobian of the non-stochastic transition in capital can be computed using (3.48). Since the integral in (3.50) is defined over the spaces of $\breve{A}_{t-1}$ and $\breve{V}_{t-1}$, numerical quadrature using the parameters of the marginal distributions of $\breve{A}_{t-1}$ and $\breve{V}_{t-1}$ from $g(s_{t-1}|\widehat{a}_{t-1})$ (two-dimensional Gauss-Hermite quadrature) was used in its

computation. The use of numerical quadrature was found to be sufficiently accurate since the Jacobian was almost always a "plane" in the two-dimensional space of $(\breve{A}_{t-1}, \breve{V}_{t-1})$. An example of the computed Jacobian is illustrated in Figure 3.5.

3.5.2.3 Likelihood Evaluation

The general approach to creating the meta-Gaussian sampler is similar to the process utilized in Example 1 above[3]. We begin with the initialization of the marginal densities and an initial correlation matrix for the copula. The marginal density of $\breve{K}_t$ was approximated using a piecewise-continuous approximation. The marginal densities for $\breve{A}_t$ and $\breve{V}_t$ were specified to be log-normal.

Given a Gaussian approximation for the prediction density $f(s_t|X_{t-1};\theta)$, linearization of the measurement equations (illustrated in Appendix C) around the steady state values of $\breve{y}_t$ and $\breve{i}_t$ provides a Gaussian approximation to the density of the state variables conditional on current observations of $\breve{y}_t$ and $\breve{i}_t$. The mean and variance of the capital stock were utilized to define the initial set of uniformly spaced nodes in the space of $\breve{K}_t$ (the lower and upper limits were $\pm 5 \times$ *standard deviations*). The mean and variance of $\ln \breve{A}_t$ and $\ln \breve{V}_t$ were utilized to initialize their respective margins. Converting the covariance matrix of the Gaussian approximation to a correlation matrix ($\Re$) provided a natural initialization for the meta-Gaussian copula density.

Having initialized the margins, N draws from a trivariate standard normal density were transformed using the correlation matrix ($\Re$) into draws of $\widetilde{s}_t$. These draws of $\widetilde{s}_t$ were transformed into s_t by the inverse of the transformation in (3.19) and the marginal densities specified above.

Analogous to the procedure employed in Example 1, the piecewise-continuous density approximation for $\breve{K}_t$ was created by approximating the marginal density at the nodes using kernel density approximation. A Gaussian kernel with the Sheather-Jones plug-in bandwidth was utilized for this purpose. An equal probability partition of $\breve{K}_t$ was obtained by inverting the piecewise-density as outlined in DeJong et al. (2007). The mean and variance of the univariate

[3]For an extensive description of the particle filter and its application to evaluate the likelihood function, refer to DeJong et al. (2007).

margins for $\breve{A}_t$ and $\breve{V}_t$ were updated using the least squares regression outlined in Appendix B.

Having updated the parameters of the margin, the correlation matrix was updated by a trivariate analogue of the process outlined in Appendix C. That is, the least-squares problem arises due to the approximation

$$\ln f(x_t|s_t;\theta)f(s_t|X_{t-1};\theta) - \sum_{j=1}^{3} \ln f(x_{t,j};\theta_j) \propto -\frac{1}{2}\left(\widetilde{s}_t'\Re^{-1}\widetilde{s}_t\right). \tag{3.51}$$

3.5.2.4 Initializing the Copula Sampler

Initializing the sampling distribution prior to the recursive EIS procedure is a non-trivial task. Simple methods like a grid-search are too expensive to implement over multiple dimensions. Building upon an approach currently being developed by DeJong et al., we create an approximate joint distribution for the observable and state variables by linearizing the measurement equation and combining it with a Gaussian approximation for the prediction density of the state variables. A Gaussian density for the state variables conditional on current information is obtained from the approximate joint distribution. Further details are provided in Appendix D.

3.5.3 Results and Discussion

The time-0 distributions of $\breve{A}_0$ and $\breve{V}_0$ were specified to be normal centered at 1 and with a standard deviation of 0.06. The distribution of $\breve{K}_0$ was set to $N(\overline{\breve{K}},(0.1)^2)$, where $\overline{\breve{K}}$ is the steady state capital stock computed with the parameters used to simulate the data generating process. A data set containing two hundred observations was simulated in order to perform maximum likelihood analysis. Further, ten data sets with each containing a single sequence of 10 draws of the latent process and 10 sequences of observables obtained from the single sequence of latent variables were also generated. The parameter values used to generate the data were, $\alpha = 0.33$, $\beta = 0.96$, $\gamma = 0.001$, $\nu = 0.002$, $\phi = 2.5$, $d = 0.1$, $\sigma_a = 0.01$, $\sigma_v = 0.02$, $\sigma_y = 0.06$, $\sigma_i = 0.04$. An example of the simulated data is shown in Figure 3.2.

Table 3.1 presents the parameter estimates results from maximizing the likelihood function using fifty and two hundred observations. The parameter estimates obtained using the smaller

the data set are relatively less accurate with the estimates of parameters like σ_a, σ_y, and σ_i changing substantially. Also, the asymptotic standard errors associated with the measurement error and the volatility of the shocks are relatively large. As the sample size increases, the ML estimates of the parameters do indeed get closer to the data generating parameters.

In order to compare the performance of the standard particle filter and the copula based EIS-Filter, two studies were conducted. Firstly, accuracy of the two procedures in computing the likelihood is compared by repeatedly computing the likelihood function for a given set of parameter values. Table 3.2 presents results obtained by computing the likelihood function using a data set containing fifty observations. Importantly, the reported times do not include the time needed for the polynomial approximation of the consumption function as this cost is common to both methods. Reiterating the results in DeJong et al. (2007) and elsewhere, computation time scales linearly with the number of draws in the case of the particle filter. However, with the EIS-Filter, we are required to choose three different numbers: the number of draws (or quadrature points) to be used to compute the prediction density (3.10); the number of draws used to construct the copula sampler for the period-t likelihood (3.8) and the number of MC draws to compute the period-t likelihood. The first of these numbers was fixed at 121 (11×11), and the second was fixed at 300. Hence the computational times of the EIS-Filer does not scale linearly with the number of MC draws used to evaluate the integral.

As Table 3.2 indicates, the EIS-Filter is substantially more accurate than the particle filter. Comparing performance using the "Relative Time Efficiency" ratio measure introduced in DeJong et al. (2007), we see that the value of this ratio varies between 4 and 25 in favor of the EIS-Filter. For example, EIS-Filter with 200 MC draws has a time-efficiency ratio of 6.8 with respect to the particle filter with 80000 particles. That is, in order to achieve the accuracy of the EIS-Filter using 100 MC draws, the particle filter would require nearly 300,000 particles. This result is reiterated by Figure 3.3 which shows the standard deviation of individual period-t likelihoods. Even with an increasing number of particles, large standard deviations continue to persist with the particle filter. In contrast, however, the standard deviations of individual likelihoods obtained using EIS-Filter are substantially more accurate even with 100

MC draws. This reflects the fact that EIS-Filter is not susceptible to problems like sample impoverishment typically associated with the particle filter.

In the second exercise, the filtering accuracy of particle filter was compared with that of the EIS-Filter using copula densities. This exercise extends the experimental design used by Pitt and Shephard (1999). For this purpose, ten different trajectories of the latent process ($\breve{A}_t, \breve{V}_t$ and $\breve{K}_t$) were obtained with each trajectory containing ten data points. From each trajectory, ten different sequences of the observable variables ($\breve{y}_t, \breve{i}_t$) were obtained, creating a total of 100 data sets. With each data set, a "pseudo-true" estimate of the filtered values of the latent variables were obtained as the mean of 20 replications of the EIS-Filter with 10000 MC draws.

Note that the Pitt and Shephard (1999) experiment involved the simulation of mutiple sequences of the observable variables from a simgle sequence of latent variables. In contrast, we obtain sequences of observables from ten different realizations of the latent variables. This is analogous to repeating the Pitt and Shephard (1999) exercise ten times to mitigate any peculiarities in the results due to particular realizations of random variables. The sequence of steps in this experiment were as follows. For each sequence of observable variables,

1. The estimated "pseudo-true" filtered values of the latent variables are computed.
2. Twenty different sets of filtered values of the latent variable are obtained using the particle filter (with 80000 particles) or the EIS-Filter (with 300 MC draws).
3. The Mean Squared Error (MSE) of the filtered values are computed by averaging over the squared difference between the above twenty estimates and the "pseudo-true" values.

Denoting $\breve{\zeta}_t^i$ as the estimated "pseudo-true" filtered value for data set i, and $\overline{\zeta}_{t,k}^{i,j}$ denoting the MC estimate of the filtered latent variable for data set i, replication j, time t and method k (EIS-Filter or Particle Filter), The Mean Squared Error (MSE) is given by

$$MSE_{t,k} = \frac{1}{100}\sum_{i=1}^{100}\left[\frac{1}{20}\sum_{j=1}^{20}\left(\overline{\zeta}_{t,k}^{i,j} - \breve{\zeta}_t^i\right)\right]. \tag{3.52}$$

Indeed, the EIS-Filter produces smaller mean squared errors. Aggregating over observations drawn from different realizations of latent variables ensures that these results are robust to any peculiarities in the results due to particular realizations of random variables. As seen

in Figure 3.4, the average difference (in log-scale) of the MSE values are around 1.0 for $\breve{A}_t$, while the average difference for $\breve{K}_t$ and $\breve{V}_t$ are nearly 2.0 and 2.5 respectively. Among the latent variables, the Log-MSE values for $\breve{K}_t$ are relatively higher than those of the rest.

3.6 CONCLUSION

This paper illustrates the use of copula densities (specifically, the meta-Gaussian copula) as importance samplers for likelihood evaluation and filtering in general state-space models. The example presented shows that the copula based importance Indeed, copula densities facilitate the construction of importance samplers with widely differing marginal distributions. The likelihood function evaluated using copula based EIS-Filtering methods possess critical properties such as, continuity with respect to underlying parameters and smoothness. These properties successfully address the twin concerns of computational cost (of accuracy) and numerical variance of the likelihood function that are significant with the use of the standard particle filter (Fernandez-Villaverde and Rubio-Ramirez, 2007).

3.7 TABLES AND FIGURES

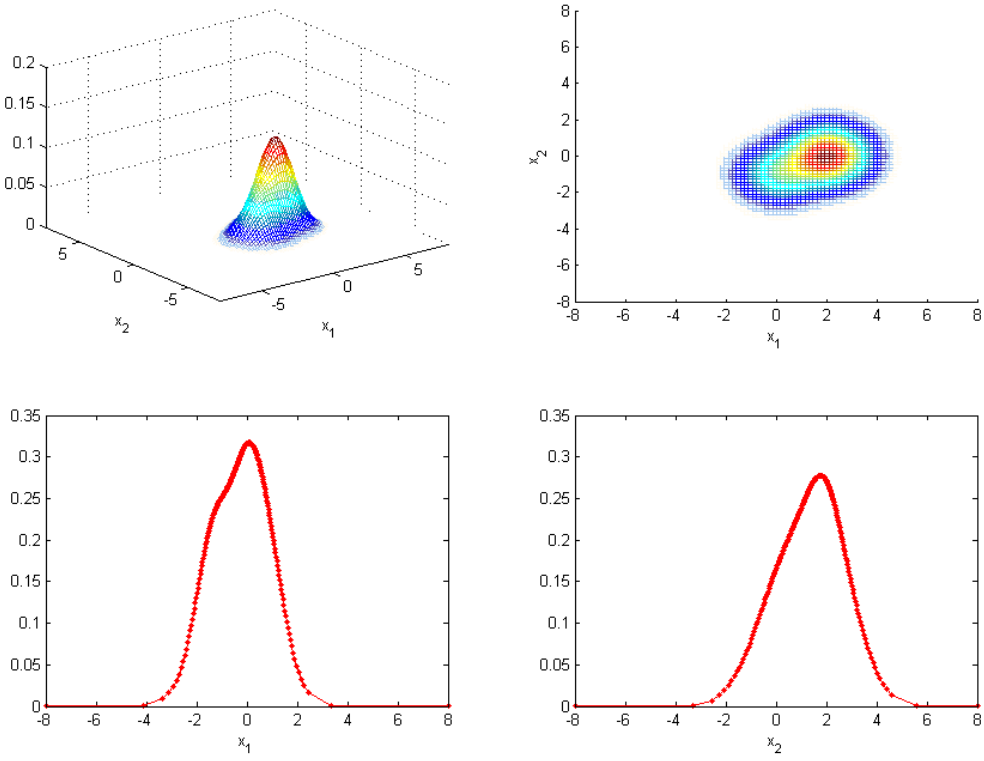
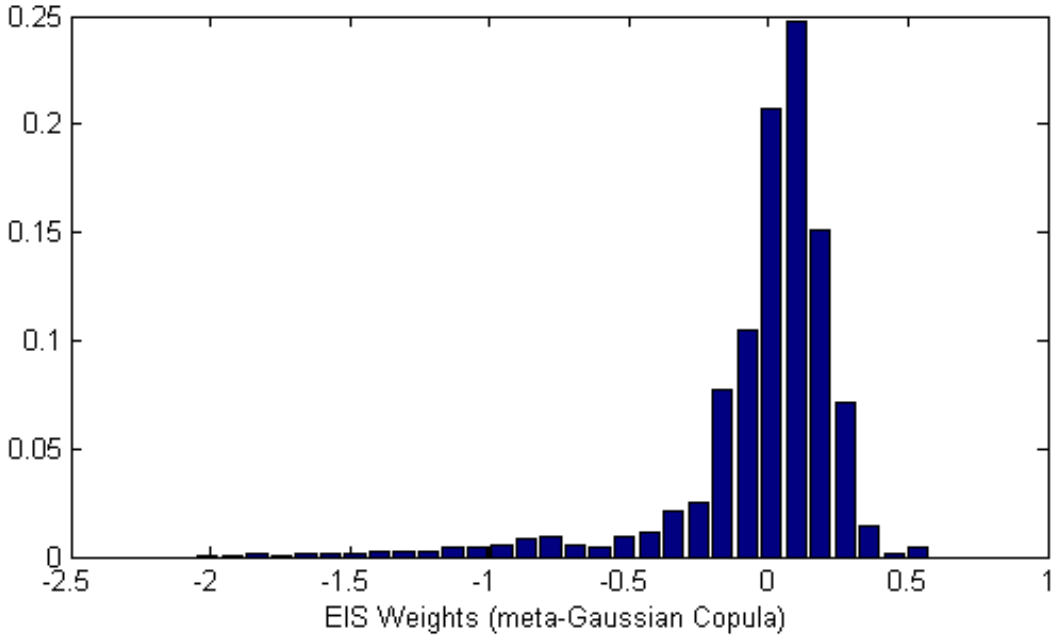


Figure 3.1: Mixture of Two Bivariate Normals (Specification-I)



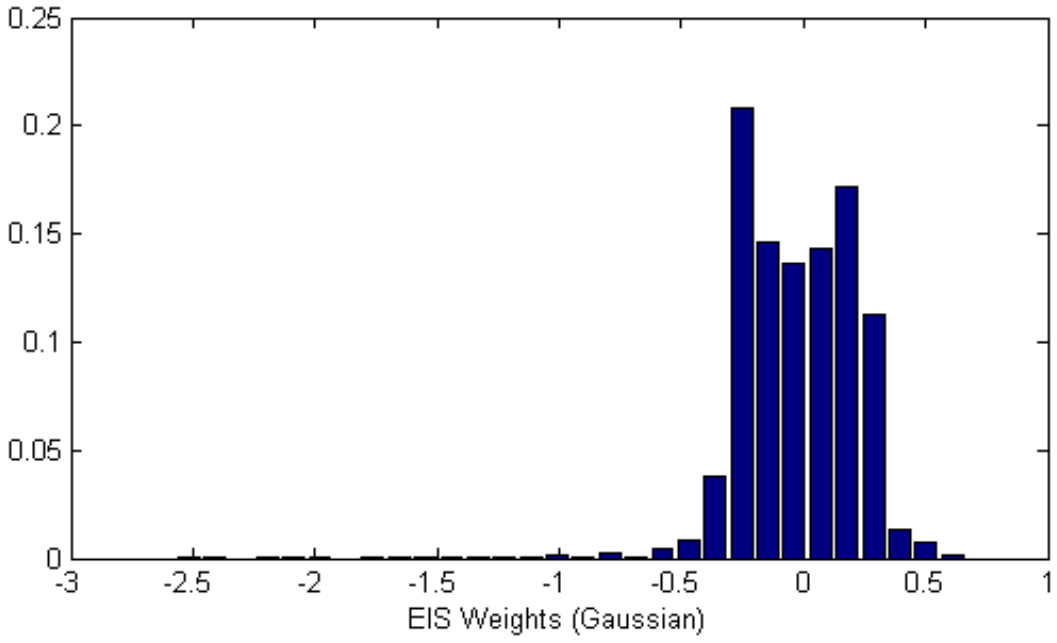


Figure 3.2: EIS Weights (Specification-I)

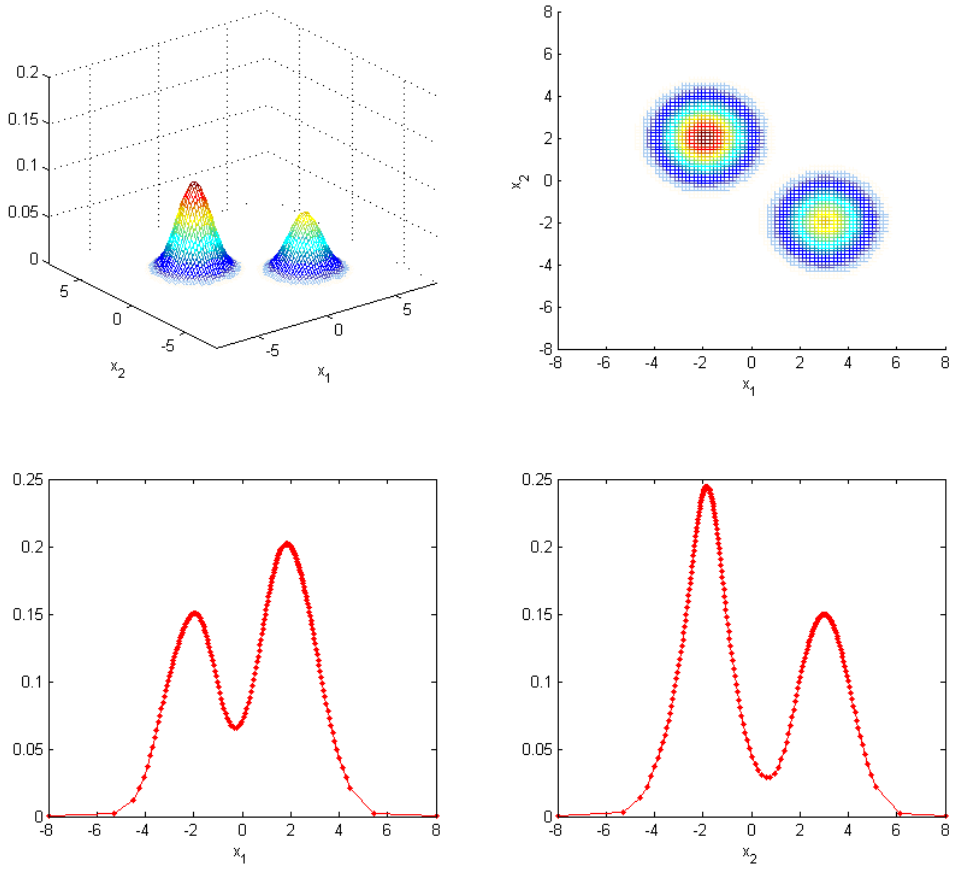


Figure 3.3: Mixture of Two Bivariate Normals (Specification-II)

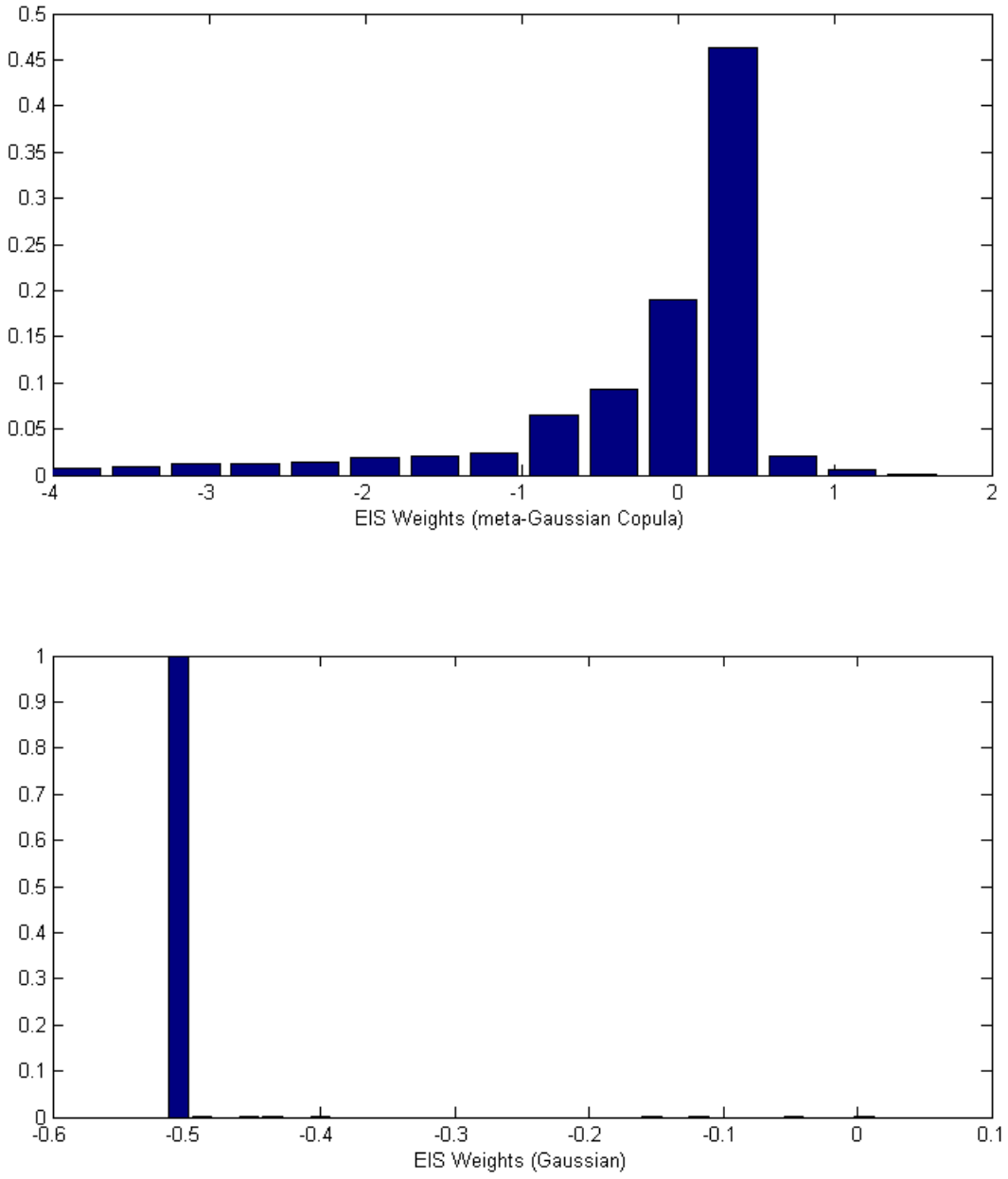


Figure 3.4: EIS Weights (Specification-II)

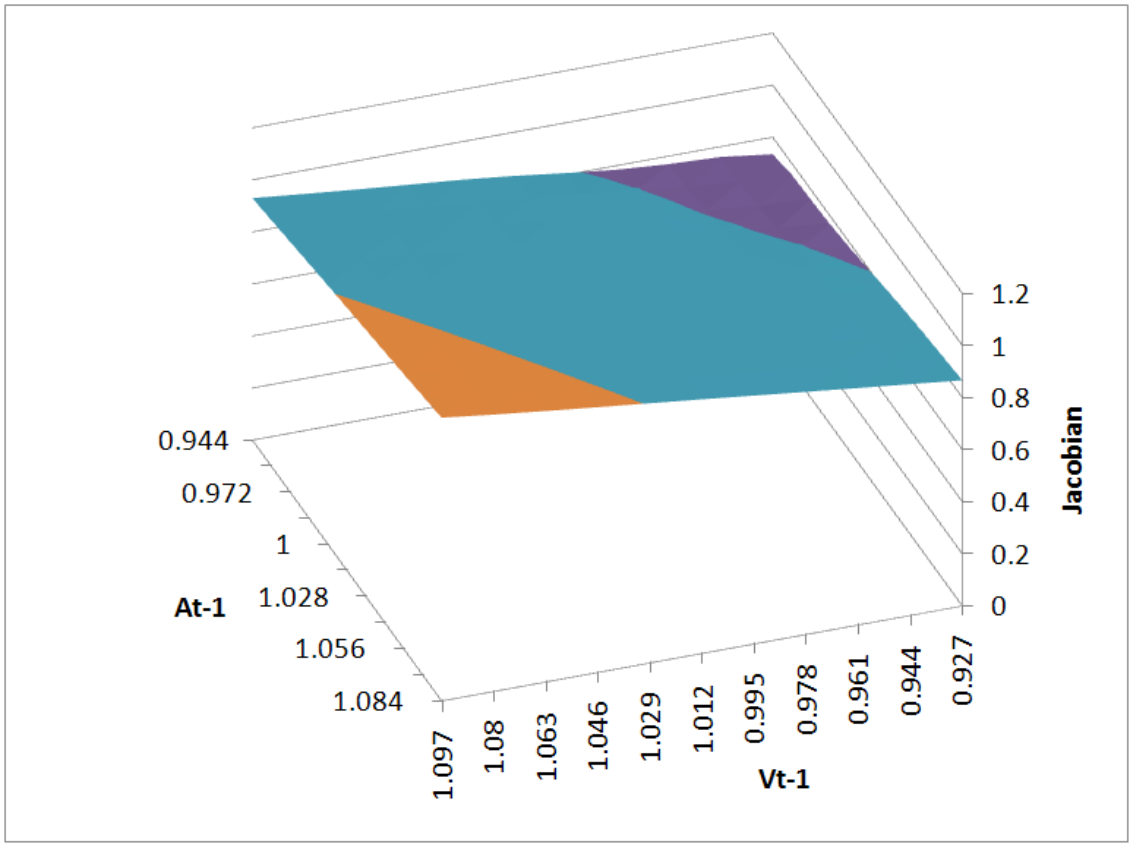


Figure 3.5: Jacobian of the Dirac transformation in $\check{K}_t$.

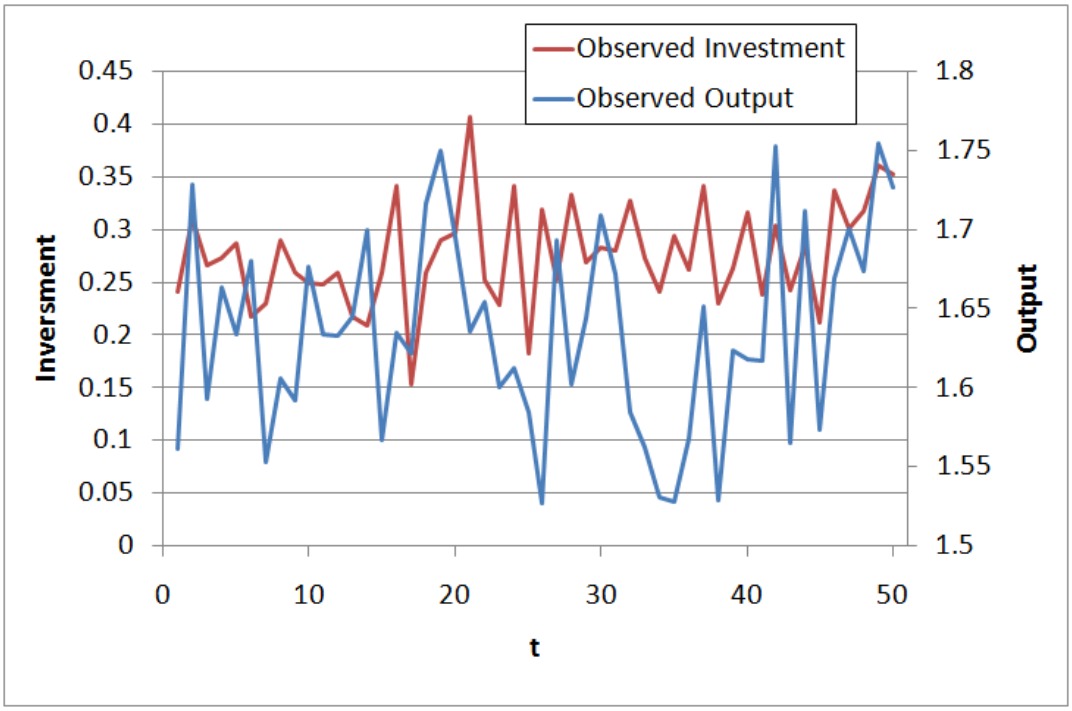


Figure 3.6: Simulated Data.

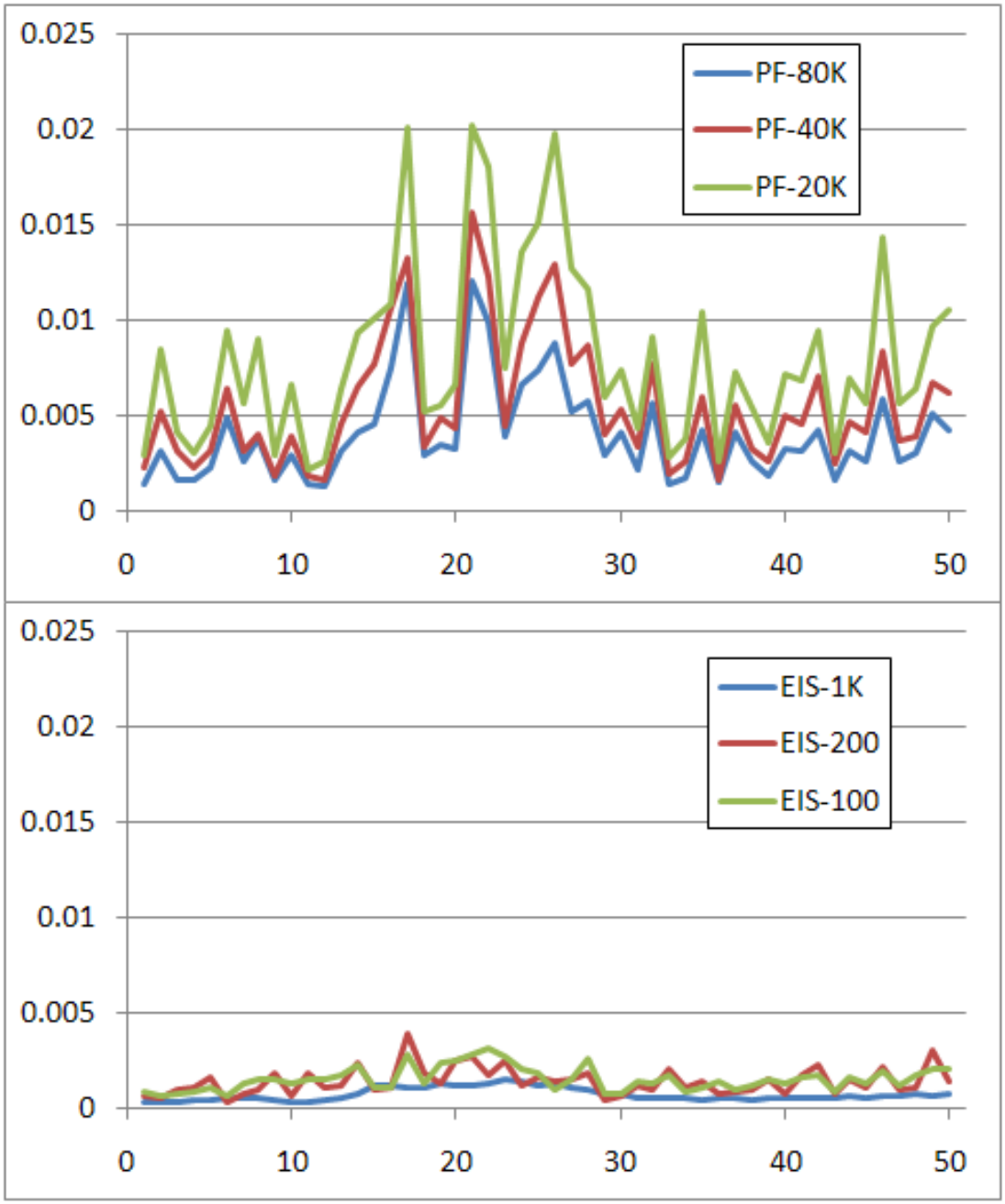


Figure 3.7: Accuracy of individual likelihoods.

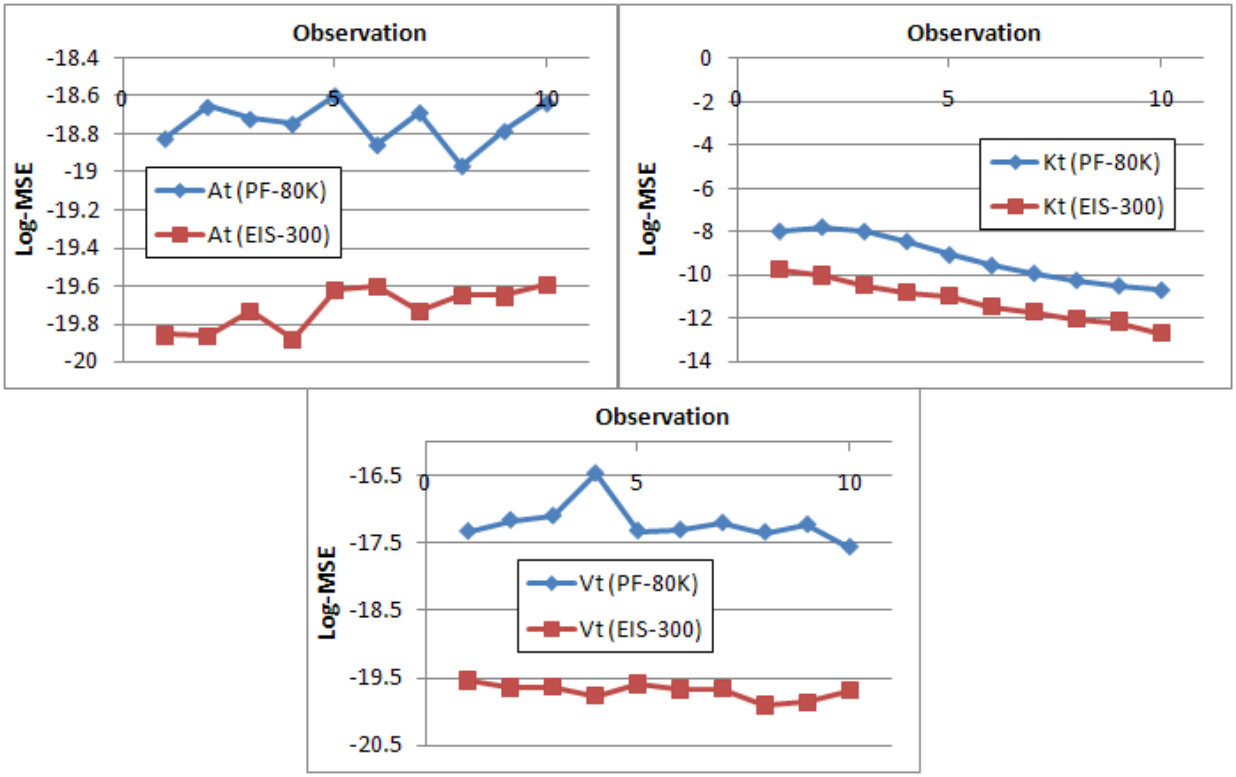


Figure 3.8: Log-MSE Comparisons.

Table 3.1: Mixture of Two Bivariate Normals

	Gaussian EIS			**meta-Gaussian Copula**		
Specification-I						
N	Mean	Stdv	Time	Mean	Stdv	Time
100	0.9921	0.0246	0.0574	1.0169	0.0293	0.2041
200	0.9955	0.0159	0.0613	1.0104	0.0205	0.2131
500	0.9980	0.0120	0.0750	1.0086	0.0106	0.2192
1000	0.9993	0.0076	0.0959	1.0062	0.0083	0.2336
5000	0.9997	0.0031	0.2555	1.0061	0.0035	0.3658
Specification-II						
	Mean	Stdv	Time	Mean	Stdv	Time
100	0.8467	1.6601	0.1027	1.0361	0.1191	0.4972
200	0.7000	0.2904	0.1069	1.0161	0.0678	0.5016
500	0.6894	0.2103	0.1238	0.9925	0.0523	0.5209
1000	0.6909	0.2040	0.1520	1.0087	0.0339	0.6342
5000	0.6809	0.1737	0.4091	0.9942	0.0299	0.8539

Table 3.2: Maximum Likelihood Estimation

	200 Observations		50 Observations	
	Parameter Estimate	**Std. Error**	**Parameter Estimate**	**Std. Error**
α	0.3334	7.769E-03	0.3431	8.314E-03
β	0.9526	1.147E-02	0.9622	1.701E-02
ϕ	2.505	1.482E-03	2.5007	2.276E-03
d	0.1802	2.371E-02	0.1857	6.808E-02
γ	0.0012	7.639E-03	0.00076	1.877E-02
ν	0.0025	7.147E-03	0.0019	1.748E-02
σ_a	0.0115	1.140E-02	0.0247	4.604E-02
σ_ν	0.0406	7.956E-03	0.0497	1.056E-02
σ_y	0.0614	9.079E-03	0.0970	1.727E-02
σ_i	0.0386	8.547E-03	0.01945	2.071E-02

Table 3.3: Computation Times

	Particle Filter				EIS-Filter		
N	**Mean**	**Stdv**	**Time(s)**	**N**	**Mean**	**Stdv**	**Time(s)**
10000	124.5688	0.0931	0.935	100	124.5843	0.01763	6.70
20000	124.5855	0.0703	1.848	300	124.5844	0.01281	7.91
40000	124.5843	0.0480	3.919	1000	124.5845	0.00443	17.59
80000	124.594	0.0330	8.189				

4.0 PREDICTING REGIME SHIFTS IN U.S. GDP GROWTH

4.1 INTRODUCTION

Success in forecasting GDP growth depends critically on the ability to anticipate shifts of the economy between episodes of general expansion and contraction. The apparent non-linearity of the process that governs these shifts, coupled with the decrease in the volatility of GDP growth observed since the mid-1980s, renders the task of anticipating regime shifts as challenging. The central aim of this paper is to build a model that can accurately detect regime-shifts in the U.S. business cycle and forecast both growth rates and potential regime-shifts.

As noted by Pagan (1999) the study of business cycle phase changes continues to be an important issue in macroeconomic research. Three concerns that emerge repeatedly in business cycle research are, (a)asymmetries in the mean-growth trajectory across regimes, (b)asymmetry in the volatility of growth rates across regimes, and (c)the importance of non-linearities in predicting turning points and capturing observed features of business cycles.

Many non-linear models attempt to capture the asymmetries in the behavior of GDP growth across regimes. For example, the Beaudry and Koop (1993) model allows for a "current depth-of-recession" variable to be triggered by negative growth rates, leading to a quick uptick of growth rates after the end of a recession (bounce-back effects). Potter (1995) presents a self-exciting threshold auto-regression (SETAR) model with two states. A negative growth in the two preceding periods implied a switch between the two states. The model by Pesaran and Potter (1997) incorporates a "floor" and a "ceiling" for growth rates: i.e., the growth rate accelerates (decelerates) when it reaches floor (ceiling).

In his seminal paper, Hamilton (1989) captures asymmetry in the mean of growth in business cycle regimes using a two-state Markov-switching model of output. An early study of conditional asymmetry in the variance of GNP growth was performed by French and Sichel (1993). They model the conditional mean and variance of U.S. GNP using an asymmetric exponential GARCH specification and find that the conditional variance is typically the largest around business cycle troughs. An extensive review and comparison of the ability of linear and non-linear models to account for stylized facts regarding business cycle regimes can be found in Hess and Iwata (1997b). They conclude that non-linearities confer little advantage over linear models in reproducing the salient features of business cycle regimes observed in data.

Also, there has been a dramatic reduction in the volatility of U.S. GDP growth since mid 1980s. This phenomenon is well documented in literature (references include Kim and Nelson 1999b, McConnell and Perez-Quiros 2000, Stock and Watson 2002 etc.) and is usually refered to as the "Great Moderation". Stock and Watson (2002) note two alternative perspectives regarding the nature of volatiloty reduction; (a) an ongoing trend (Blanchard and Simon 2001) and (b) a sharp break as suggested by Kim and Nelson (1999b).

Building on the work of DeJong, Liesenfeld, and Richard (2005) we develop a model (henceforth the DHLR model) designed to allow for asymmetry in both mean-trajectory of growth and its variance. The model characterizes GDP growth as following non-linear trajectories that fluctuate stochastically between alternating episodes of general acceleration and deceleration. Regime changes occur stochastically, with probabilities determined by an observed indicator variable via a logistic link function. We refer to the indicator variable as a "tension index" , which is constructed as a geometric sum of past deviations of actual GDP growth from a corresponding "sustainable rate" (interpretable as the growth rate of potential GDP).

Predicting regime-shifts is, in general, a difficult problem. In the context of the DeJong et al. (2005) and DHLR models, detection of a regime-shift critically depends on both the logistic link function and the ability of the non-linear drift process to detect subtle shifts in the

general direction of GDP growth. The proposed (DHLR) model is a substantial improvement over the DeJong et al. (2005) model with respect to the latter issue. With the observed decline in volatility of growth, detecting changes in the general direction of growth requires substantial flexibility in the behavior of the non-linear drift process to respond to subtle changes. The motivation for the change can be traced back to the fact that the DeJong et al. (2005) model repeatedly failed to identify the current contractionary regime. In sharp contrast, the DHLR model indicates that the economy transitioned into contraction in 2003:III.

To account for the fact that no two business cycles are alike, we model the parameters that dictate the non-linear trajectories as unobservable random variables that vary across regimes. At the beginning of each regime, a new set of parameters is drawn from a fixed distribution. These parameters consist of the initial value of the level, the velocity, and the acceleration of the regime-specific non-linear trajectory (termed the "drift process"). This permits the model to capture a wide range of behavior across regimes for the drift process, from nearly linear to brief exponential bursts.

Our model allows for volatility of growth rates to differ across regimes. It is specified to be a latent variable, whose regime-specific value is drawn from a known distribution at the beginning of each regime. Thus, regimes with low volatility correspond to low realizations from the associated distribution. As noted in later sections, results from our model also suggest a sharp reduction in the volatility of GDP growth since the 1980s. However, it is difficult to explicitly account for this trend in our model specification with only 17 available observations for the volatility. While we are left with the possibility of a specification error in our model, the increased "sensitivity" of the latent-drift process implies that our model adjusts to new information very quickly. Thus, despite beginning with a conservative forecast for the volatility of growth in a new regime, the latent-process adapts quickly to new information and correct its initial forecast as required. Also, a conservative initial-estimate for the volatility protects us from making extreme assertions regarding the volatility of growth and leave the door open for possible future increases.

We estimate the model via maximum likelihood using two alternate estimation procedures.

The first conditions on a sequence of regime-change dates (the "optimal" sequence computed in an iterative procedure described in later sections); the second is unconditional on any fixed sequence of regime-change dates (results are obtained by integrating over all possible regime-change dates). We study the ability of the model to anticipate and/or detect regime-shifts and contrast its performance with that of the DeJong et al. (2005) model.

Marcellino (2008) presents an extensive comparison of the forecasting performance of univariate time series models for U.S. GDP growth and inflation. Comparing the forecasting accuracy of linear AR specifications, time-varying AR specifications, smooth transition AR specifications, SETAR, Markov-switching and other models (using mean-squared-errors of out-of-sample forecasts over a long time horizon), they report that nonlinear models confer limited gains over well-specified linear models.

As the next section indicates, our model can be expressed as an ARMA(3,1) process with a non-linear drift. In order to test the relative merits of a non-linear specification, we compare the errors in one-period-ahead out-of-sample forecasts of GDP growth obtained from the non-linear model to those from a linearized version (where the non-linear drift process is replaced by a linear one). We also compare forecasting accuracy of our model with that of DeJong et al. (2005) and random-walk-with-drift specifications. Importantly, the DHLR model and the DeJong et al. (2005) have lower forecast errors relative to the linearized-DHLR model; highlighting the added benefit of non-linearities in forecasting growth. The DHLR and DeJong et al. (2005) models perform similarly during the long regimes. However, the DHLR model is more accurate (lower MSE) in short regimes. Also, the DHLR model outperforms the random-walk specification in all the regimes and in the overall comparison.

In addition to detecting changes in the general direction of GDP growth, forecasts of growth from our model can also be utilized to anticipate the occurrence of recessions. We propose a non-stochastic recession-dating rule that utilizes growth rates and changes in the tension index to determine if a recession has begun, and (conditional on a recession having begun), it if will end at a particular quarter.

The most widely accepted dating of recessions is that done by NBER. As noted by Chau-

vet and Hamilton (2005) and Chauvet and Piger (2008), these dates are determined by the NBER's "Business Cycle Dating Committee" using a wide variety of economic indicators and represent the consensus opinion of many individuals. Hence, they conclude that the NBER dating method is neither transparent nor reproducible. While economic data are periodically revised, the NBER announced dates have seen little revision. Often, these dates are announced after a substantial delay (e.g., March 1991 and November 2001 business cycle troughs were announced nearly two years after the event).

Both statistical and deterministic alternatives to the NBER dating methods can be found in literature. Bry and Boschan (1971) show that it is possible "mimic" the NBER committee's judgements through a formal algorithm. Harding and Pagan (2002a) note the existence of different approaches to determine turning points in the business cycle. In the context of univariate models they classify them as parametric and non-parametric methods. The non-parametric methods include the Bry and Boschan (1971) and other rule-based approaches which eschew a formal statistical model. The parametric class includes methods that utilize parameters estimated from a formal statistical model to determine turning points (e.g. Markov-switching model of Hamilton 1989).

In a study comparing the quarterly analogue of the Bry-Boschan non-parametric method (BBQ) and Markov-Switching models, Harding and Pagan (2002a) show that while both methods perform reasonably well, recession-dating with MS models is sensitive to actual specification and the choice of critical values (in probability weights) utilized to make decisions regarding turning points. Arguing in favor of the BBQ method they highlight its robustness to data generating processes and the sample selected.

As seen in Figure 4.1, the tension index serves as a good barometer for recessions. Our recession-dating method a hybrid between the parametric and non-parametric methods. The simplicity of a rule-based recession-dating method makes it an attractive alternative to parametric probability models. For the purposes of forecasting, trajectories of growth are simulated using the DHLR model as a data generating process. For each simulated trajectory, the "optimal" quarter for the start of a recession is computed using the dating rule. Conditional

on recession having started at the specific quarter, its terminal quarter is also determined via the dating rule.

The results indicate that the DHLR model is capable of exhibiting substantially more non-linear behavior in its regime-specific latent process and represents a major improvement over the DeJong et al. (2005) model in its ability to anticipate and/or detect regime-shifts. The current model also has lower forecast errors relative to the linearized version, the DeJong et al. (2005), and the random-walk models. Forecasts of probability of regime-shifts, growth rates, and probability of a recession in the near future all highlight the importance of increased non-linearities of the proposed model.

4.2 THE MODEL

Our analysis builds upon on the non-linear forecasting model discussed in DeJong et al. (2005). The mean specification is common to both versions, and has two key features: an error correction mechanism (ECM) and regime-switching behavior. Regime switches are manifested in the behavior of a stochastic drift component that fluctuates between periods of general acceleration and deceleration. Regime changes are triggered by a tension index, constructed as a geometric sum of past deviations of actual GDP growth g_t from a corresponding "sustainable" growth rate (interpreted as the growth rate of potential GDP). Letting g_t^* denote the sustainable rate, and y_t the deviations $y_t = g_t - g_t^*$, the tension index G_t is given by

$$G_t = \sum_{i=0}^{\infty} \delta^i y_{t-i} \,, \tag{4.1}$$

where $0 < \delta < 1$ measures the persistence of past deviations on current G_t. Since g_t^* is unobservable, we specified g_t^* as the sample mean of g_t. By implication, g_t tends to pass between phases during which it alternately tends to outstrip and fall short of g_t^*. Under the interpretation of our model, neither phase is sustainable: both produce tension buildups (captured by increases in the absolute value of G_t) that ultimately lead to regime changes. We model

regime-change probabilities using the following logit specification:

$$\pi_t = P(s_{t+1} = -s_t | s_t, G_t) = \frac{1}{1+\exp\{\beta_0 - \beta_1 s_t G_t\}} , \tag{4.2}$$

where the variable s_t indicates the regime prevailing in period t, being 1 if G_t is in an expansionary regime and -1 otherwise. Thus as the absolute value of G_t increases, so too does the probability of a regime change.

The model for GDP growth, in terms of its deviations from g_t^*, is given by

$$y_t = m_t - \nu G_{t-2} + \gamma y_{t-2} + \varepsilon_t , \quad \varepsilon_t | \sigma_t \sim N(0, \sigma_t^2) , \tag{4.3}$$

where m_t represents a stochastic latent regime drift (the rationale behind this lag-2 specification follows the presentation of equations (4.5) and (4.6) below). Subtracting y_{t-2} from both sides of (4.3) casts the model explicitly as an ECM representation, in which the term νG_{t-2} reflects an integral correction based upon cumulated past deviations from equilibrium, and $(1-\gamma)y_{t-2}$ represents a proportional correction following the terminology of Phillips (1954 and 1957).

The specification for the regime drift, which allows for jumps at dates featuring a regime change and for different trajectories across the regimes, has the form

$$m_t = \overline{m}_j + s_t a_j \left(\frac{e^{\tau b_j} - 1}{b_j} \right) , \quad \tau \in \left\{ 0, 1, \ldots \left(t_{(j)} - t_{(j-1)} - 1 \right) \right\} , \tag{4.4}$$

where the index j ($j : 1 \rightarrow J$) denotes the regime prevailing in period t, and $t(j)$ denotes the date at which j gives way to regime $j+1$ (i.e. $t(j)$ is the last period under regime j). The variable $\overline{m}_j$ represents the value of the regime drift in the first period of regime j, and the exponential term dictates the curvature of the m_t trajectory during regime j. Specifically, a_j represents the velocity $\left(\left[\frac{dm_t}{d\tau} \right]_{\tau=0} \right)$ of the drift process at $\tau = 0$, and $a_j b_j$ the acceleration $\left(\left[\frac{d^2 m_t}{d\tau^2} \right]_{\tau=0} \right)$. Both a_j and b_j are restricted to being non-negative. For very small values of b_j the ratio $\left(\frac{e^{\tau b_j} - 1}{b_j} \right)$ is approximated using a second-order Taylor series expansion.

In contrast, the specification of m_t in DeJong et al. (2005) model implies zero initial velocity and acceleration. Of the two parameters governing the rate of change of m_t (in DeJong et al. (2005) model), one is regime-specific and the other (denoted "d") is common to all the

regimes. The DHLR model, however, allows for the initial velocities and accelerations of m_t to change across regimes. Also, all parameters governing the curvature of m_t can differ across regimes. These features provide the critical flexibility required by the non-linear drift to respond to subtle shifts in the general direction of GDP growth.

Pre-multiplying by $(1-\delta L)$, (4.3) can usefully be rewritten in the form of an (overidentified) ARMA(3,1) plus drift process

$$y_t = n_t + \delta y_{t-1} + (\gamma - \nu) y_{t-2} - \gamma\delta y_{t-3} + \varepsilon_t - \delta\varepsilon_{t-1} \ , \tag{4.5}$$

where within regime j the variable n_t is given by

$$n_t = (1-\delta)\overline{m}_j + \frac{a_j}{b_j} s_t \left[e^{\tau b_j} - \delta e^{(\tau-1)b_j} - (1-\delta) \right] \ . \tag{4.6}$$

Its overidentification provides significant efficiency gains in the estimation stage at virtually no loss of fit. Note that the selection of lag 2 for the ECM representation (4.3) allows us to parsimoniously capture a non-zero coefficient on y_{t-3} in (4.5), which turns out to be statistically significant.

In order to allow for variation in the drift proces m_t across the J regimes, we assume that a_j, b_j and $\overline{m}_j$ are latent random variables. An extensive diagnostic analysis led to the specification of a trivariate normal distribution for $\ln a_j$, $\ln b_j$ and $\overline{m}_j$ with their mean vector and covariance matrix treated as model parameters.

Given the fact that business cycle durations are highly variable, and can be either very short (e.g., the double-dip recessions of 1979/1981) or very long (e.g., the extended expansion of the 1990s), the drift process must be flexible and capable of exhibiting non-linear behavior. As illustrated in Figure 4.2, (computational details follow) the drift process in some short-duration regimes (e.g. 1954:I-1955:I and 2000:I-2001:III) is almost linear, while in other short regimes (e.g. 1978:III-1980:II and 2001:IV-2003:III), it exhibits highly non-linear behavior. Likewise, long-duration regimes like the 1984:II-1991:I regime exhibit pronounced acceleration towards the end. However, other long-duration regimes (e.g. 1991:II-1999:IV) require much smaller acceleration. Also, Figure 4.2 shows that all long-duration regimes exhibit a drift process that grows very slowly before the "jump" as the regime nears it end. The top-right panel

of Figure 4.2 illustrates the drift process since the mid-1980s (corresponding to the Great Moderation period), highlighting the better fit of the new drift process relative to the drift process in DeJong et al. (2005).

Finally, to capture the observed heterogeneity in GDP volatility, particularly with the Great Moderation, we specify that the conditional variance of growth-rate innovations σ_t^2 is a log-normal random variable whose mean and variance are treated as model parameters. Combined with the distribution of $\ln a_j$, $\ln b_j$, and $\overline{m}_j$, the density of the four latent regime-drift parameters is written as

$$\left[\ln a_j, \ln b_j, \overline{m}_j, \ln \sigma_j^2\right]' \sim N_4\left(\mu_{\Lambda_j}, \Sigma_{\Lambda_j}\right), \tag{4.7}$$

where Λ_j denotes the four regime-drift parameters in regime-j. We explored many alternative functional forms for the relationship between σ_t^2 and other regime-specific latent parameters, but found that they did not deliver improved performance. Note that this specification is capable of accounting for discrete jumps in the innovation variance component of the model in a way that is structurally stable (i.e., in a way that does not require the specification of variance-regime dummies). This feature is critical for forecasting, since under a dummy variable specification, an in-sample variance reduction automatically extends over the forecast horizon. In contrast, by allowing for stochastic regime changes, the DHLR model does not require that the volatility reduction be permanent: new regimes imply new variance levels.

To characterize estimation, let θ represent the vector of all model parameters, X_T the data, and $\{e_t\}_{t=1}^T$ a vector of zeros and ones, where $e_t = 1$ indicates a regime change period (i.e., the next period is the beginning of a new regime). Let $\{t_j\}_{j=1}^J$ represent the vector of regime change periods. The likelihood function can be written as

$$L(\theta; X_T) = \prod_{j=1}^{J} l_j(\theta) \tag{4.8}$$

$$l_j(\theta) = \int f(X_j, \Lambda_j; \theta) d\Lambda_j, \tag{4.9}$$

where $X_T = \{X_j\}_{j=1}^J$ represents J blocks of the data each corresponding to one regime and Λ_j represents regime-j latent variables ($\ln a_j$, $\ln b_j$, $\overline{m}_j$, $\ln \sigma_j^2$). Incorporating the distributional

assumptions, each regime-likelihood can be written as

$$l_j(\theta) = \int \frac{1}{\sigma_j^2}\phi\left(\frac{x_t-\mu_t}{\sigma_j}\right)\pi_t^{e_t}(1-\pi_t)^{(1-e_t)} f(\Lambda_j)\, d\Lambda_j, \tag{4.10}$$

where $\phi(.)$ represents the standard normal density and μ_t is given by

$$\mu_t = m_t - \nu G_{t-2} + \gamma y_{t-2}. \tag{4.11}$$

The model parameters are estimated via maximum likelihood using two alternate methods. The first method involves conditioning on the set of regime-change dates $\{t(j)\}_{j=1}^{J}$. This is accomplished using an iterative procedure which ensures the selection of a coherent set of dates. The procedure begins with the specification of an initial sequence of regime-change dates, for which we obtain conditional maximum likelihood (ML) estimates of θ. Next, the estimated model is used to assess the validity of each of the chosen dates. For a given date, this involves calculating the probability (according to the estimated model) that the date in fact featured a regime-shift, relative to a sequence of alternative possibilities. These include the possibility that the regime-shift occurred at an alternative date in a given neighborhood of the chosen date, and that no regime-shift occurred during the time period in question.

To explain how this is done, let B_0 be a J-vector containing the initial sequence of chosen dates, the jth element of which is $t(j)$. Also, let $B_{0,-j}$ denote a corresponding $(J-1)$-vector constructed by excluding the jth element of B_0. Next, let θ_0 denote the ML estimate of θ obtained using B_0. Finally, let $t_1 = (t(j-1)+1)$, and $t_2 = (t(j+1)-1)$, so that $[t_1, t_2]$ represents the complete set of dates between the $(j-1)$st and $(j+1)$st regime-shift dates. Then the probability that the jth regime-shift occurred at any point t over the range $[t_1, t_2]$ is given by

$$P(t(j)=t|B_{0,-j}) = \frac{L(\theta_0|t(j)=t)}{L(\theta_0|t(j)=0)+\sum_{s=t_1}^{t_2} L(\theta_0|t(j)=s)}, \tag{4.12}$$

where $L(\theta_0|t(j)=t)$ denotes the value of the likelihood function evaluated using θ_0, conditional on the augmentation of $B_{0,-j}$ with the additional regime-shift point $t(j)=t$, and $L(\theta_0|t(j)=0)$ denotes the value of the likelihood function obtained without augmenting $B_{0,-j}$. To assess the validity of the jth regime-shift date specified under B_0, we evaluated (13) for

each date t in the range $[t_1, t_2]$, as well as for $t = 0$. If $t = t(j)$ was assigned the highest probability it was validated; otherwise, $t(j)$ was re-aligned to coincide with the date that was assigned the highest probability. Once this process was completed for each of the J elements of B_0, we specified a second vector B_1, obtained a second set of parameter estimates θ_1, and repeated the process. The procedure ends when the dates chosen in a given round represent the most likely scenario according to the parameter estimates obtained during that round. In our experience, this process typically converged within three rounds. This iterative procedure is akin to the EM-type algorithms typically employed in estimated Markov-switching models (as outlined, e.g., in Hamilton, 1994, and Diebold, Lee, and Weinbach, 1994). The regime-shift dates identified in this manner and their associated probabilities are reported in Table 4.1.

The second estimation method integrates over all possible regime-shift scenarios and is therefore unconditional. The procedure we employ is illustrated in DeJong et al. (2005). This requires the integration of the likelihood function in (4.9) with respect to probabilities associated with the vector representing regime-shift periods. In general, this is a computationally expensive problem due to the need for evaluating the marginal likelihood over all 2^T possible realization of the sequence $\{e_t\}_{t=1}^T$. DeJong et al. (2005) implemented an importance sampling procedure wherein sequences of $\{e_t^*\}_{t=1}^T$ are drawn using the values of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ and $\{\widehat{e}_t\}_{t=1}^T$obtained from the conditional likelihood estimation. Let $h(.)$ and $q(.)$ represent a decomposition of the likelihood into a conditional density for X_t and a sequence of conditional densities for e_t. Correspondingly, let θ_h and θ_q represent a partition of the parameters associated with $h(.)$ and $q(.)$ respectively ($\theta_q = \{\beta_0, \beta_1\}$). Here

$$q\left(e_t | X_t; \theta_q\right) = \pi_t^{e_t} \left(1 - \pi_t\right)^{(1-e_t)}, \tag{4.13}$$

and $h(.)$ represents components of (4.9) and (4.10) not associated with $q(.)$. Trajectories of $\{e_t^*\}_{t=1}^T$ are simulated using the values of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ from the conditional ML estimation. That is, given the "optimal" sequence of $\{\widehat{\pi}_t\}_{t=1}^T$ from the conditional model and setting $s_1 = -1$, draws of s_{t+1} conditional on s_t are drawn using the values $\{\widehat{\pi}_t\}_{t=1}^T$ and draws from the uniform distribution $U[0,1]$. This sequence of $\{s_t\}_{t=1}^T$ is easily converted to a sequence $\{e_t\}_{t=1}^T$. It is

immediately apparent that despite the use of importance sampling, unconditional estimation is indeed a computationally expensive procedure.

Both conditional and unconditional estimation require the use of numerical integration techniques to compute the likelihood function. For example, note from (4.10) that each regime requires one integral over the four-dimensional space $(\ln a_j,\ \ln b_j,\ \overline{m}_j,\ \ln \sigma_j^2)$. To accomplish this, we utilize the efficient importance sampling (EIS) procedure introduced by Richard and Zhang (2007). The likelihood function is maximized using the Nelder and Mead (1965) Simplex algorithm implemented in FORTRAN (using the AMOEBA algorithm of Press, Teukolsky, Vetterling, and Flannery, 1996).

The next subsection provides a brief description of the implementation of EIS in our problem, followed by a discussion of the results.

4.2.1 EIS and numerical integration of the likelihood function

The objective of EIS is to create importance sampling densities that are continuous, smooth and provide close approximations of the integrand, thus providing integrals with high numerical accuracy. The key efficiency properties of EIS are obtained due to the process of iteratively "refining" the parameters of the sampling density. A detailed description and comprehensive examples can be found in Richard and Zhang (2007). Let the integral shown in (4.10) be denoted as $\varphi(x_t;\theta,\Lambda_j)$, i.e.,

$$l_j(\theta) = \int \varphi(x_t;\theta,\Lambda_j)d\Lambda_j. \tag{4.14}$$

The objective of importance sampling is to introduce a sampling density $g(\Lambda_j;\alpha_g)$ (α_g represent the parameters of the sampling density and is commonly referred to as the "auxiliary parameter") from which it is much easier to obtain draws. The integral then can be written as

$$l_j(\theta) = \int \frac{\varphi(x_t;\theta,\Lambda_j)}{g(\Lambda_j;\alpha_g)} g(\Lambda_j;\alpha_g)d\Lambda_j. \tag{4.15}$$

Let $k(\Lambda_j;\alpha_g)$ denote the kernel of the density $g(\Lambda_j;\alpha_g)$ and $\chi(\alpha_g)$ the integrating constant, i.e., $\chi(\alpha_g) = \int k(\Lambda_j;\alpha_g)d\Lambda_j$. The EIS procedure of Richard and Zhang (2007) obtains the

optimal sampler parameters $\widehat{\alpha}_g$, which minimizes the variance of the ratio $\frac{\varphi(x_t;\theta,\Lambda_j)f(\Lambda_j)}{g(\Lambda_j;\alpha_g)}$. That is, the optimal sampler parameters are a solution to the minimization problem

$$\widehat{\alpha}_g, \widehat{c}_g = \arg\min_{\alpha_g, c_g} \int \left[\ln \varphi(x_t;\theta,\Lambda_j) - c_g - \ln k(\Lambda_j;\alpha_g)\right]^2 g(\Lambda_j;\alpha_g) d\Lambda_j, \tag{4.16}$$

where c_g is an intercept meant to calibrate $\ln\left(\frac{\varphi(x_t;\theta,\Lambda_j)}{g(\Lambda_j;\alpha_g)}\right)$. Since the auxiliary parameters α_g are a part of the sampling density $g(.)$, the above problem is computationally implemented as a fixed-point problem. The computational steps can be summarized as follows.

1. Given a current estimate $\{\alpha_g, c_g\}^n$ of the auxiliary parameters (the process begins with the initialization of $g(\Lambda_j;\alpha_g)$ by choosing values of α_g), R values of the latent variables are drawn from $g(\Lambda_j;\alpha_g^n)$.
2. Updated values of $\{\alpha_g, c_g\}$ are obtained as the solution to the problem

$$\{\alpha_g, c_g\}^{n+1} = \arg\min_{\alpha_g, c_g} \sum_{i=1}^{R} \left[\ln \varphi(x_t;\theta,\Lambda_j^i) - c_g - \ln k(\Lambda_j^i;\alpha_g)\right]^2. \tag{4.17}$$

3. The above steps are repeated until convergence in the values of the sampler parameters is reached.

To obtain fast convergence, and to ensure continuity of corresponding likelihood estimates, draws of $\left\{\Lambda_j^i\right\}_{i=1}^R$ are obtained by transforming a set of common random numbers (CRNs) drawn from a canonical distribution (i.e., one that does not depend on α_g). When convergence is reached, the integral is given by

$$\widehat{l}_j(\theta) = \frac{1}{N}\sum_{i=1}^{N} \omega_i, \tag{4.18}$$

$$\omega_i = \frac{\varphi(x_t;\theta,\Lambda_j^i)}{g(\Lambda_j^i;\widehat{\alpha}_g)}. \tag{4.19}$$

When the sampling density is Gaussian, the minimization problem in (4.17) reduces to an ordinary least squares problem. The above process can also be used to obtain smoothed values of the latent variables. That is,

$$E(\Lambda_j|X_t,\theta) = \frac{\int \Lambda_j \varphi(x_t;\theta,\Lambda_j) d\Lambda_j}{\int \varphi(x_t;\theta,\Lambda_j) d\Lambda_j}. \tag{4.20}$$

Typically, draws from the EIS sampler $g(\Lambda_j;\widehat{\alpha}_g)$ are sufficient to compute the smoothed values of the latent variables, i.e.,

$$\widehat{E}(\Lambda_j|X_j,\theta) = \frac{\sum_{i=1}^{N} \Lambda_j \omega_i}{\sum_{n=1}^{N} \omega_n}. \tag{4.21}$$

4.2.2 A Deterministic Procedure for Dating Recessions

Figure 4.1 shows that the local troughs in the tension index (G_t) tend to (mostly) coincide with NBER-defined business cycle troughs. Thus the tension index appears to be a fairly reliable indicator for NBER-defined recessions. Forecasts of growth from our model can also be utilized to anticipate the occurrence of NBER-defined recessions. We propose a non-stochastic recession-dating rule that utilizes growth rates and changes in the tension index to determine if a recession has begun, and (conditional on a recession having begun), if it will end at a particular quarter, automatically providing us its duration.

A popular rule-of-thumb method for dating recessions is the "two quarters of negative growth" rule. An examination of quarterly growth data does show that each recession contains at least two quarters of negative growth. Figure 4.1 shows that the starts of recessions typically coincide with G_t decreasing rapidly. Also, all recessions were typified by G_t falling to a value below -5. These two observations indicate that G_t has to decrease and become sufficiently low for it to signify the possibility of a recession. In order to determine the change in G_t, it was found to be optimal to utilize the difference in local averages of G_t. Thus, for each quarter t, we compute a local change in average G_t as

$$\Delta\overline{G}_t = \overline{G}_{[t,t+3]} - \overline{G}_{[t-3,t]}, \tag{4.22}$$

where $\overline{G}_{[t,t+3]}$ represents the arithmetic mean of the tension index over the 4 period window $[t,t+3]$. The rules of our procedure are summarized as follows.

- A period t can serve as a potential candidate for the start of a recession if at least two negative growth rates are observed in a four period forward-looking window in the windows

$[t, t+3]$ are negative. For example, let t be the first period encountered with at least two negative growth rates in the windows $[t, t+3]$. Let $t+k$ be the last contiguous period to satisfy this criterion. That is, the sequence of dates $t, t+1, ..., t+k$ serve as potential starts for a recession.

- Given the sequence of potential start-dates ($[t, t+k]$), periods $t+3, t+4, ..., t+k+3$ automatically serve as potential candidates for the end of the recession.

- Having identified a contiguous sequence of periods as potential start-dates, we choose the optimal start-date according to the following criteria.

  1. Periods with the tension index larger than -5 are eliminated.
  2. Among the remaining periods, the one corresponding to minimum local change in average tension index ($\Delta\overline{G}$) is chosen as the optimal start-date. For the purposes of illustration, let period $t+j$ be the optimal start-date.

- Hence, the sequence of periods $[t+j+1, t+k+3]$ serve as potential candidates for the end of the recession. The optimal end-date for the recession is chosen as follows:

  1. Periods with positive tension index values are eliminated.
  2. Beginning with the last period in the remaining group of potential end-dates, we eliminate periods with with contemporaneous growth rate greater than 2.75 and growth rate greater than -5 in the preceding period. The first period not eliminated is the optimal choice for the end of the recession.

Application of the above dating procedure to quarterly U.S. GDP data from 1950:III to 2008:I results in near accurate prediction of the starts and ends of recession. With the starts of recessions, the algorithm was off by one quarter on three occasions (1969:III, 1979:IV and 1981:IV were declared as starts for recessions whose NBER starts were 1969:IV, 1980:I and 1981:III respectively). With the ends of recessions, the procedure was off on only one occasion: the recession with the NBER dated end in 1991:I was determined to be at 1991:III by our method.

4.3 RESULTS

4.3.1 Estimation of Model Parameters and Smoothed Latent Drift Process

The sequence of regime-shift dates used to obtain the conditional ML parameter estimates are illustrated in Table 4.1. Conditional and unconditional ML parameter estimates are reported in Table 4.2. The data set consisted of quarterly U.S. GDP data (in chained 2000 dollars) spanning third quarter of 1950 to the first quarter of 2008. A Monte Carlo sample size of 1000 was used to integrate over regime-shift dates in obtaining the unconditional estimates. Computation of the likelihood function required the evaluation (using EIS) of one integral for each regime. The size of the auxiliary regression (denoted R in the description above) was 500, and the Monte Carlo sample size used to compute the integral using the optimal sampling density was also 500.

It is important to note that the data set used by DeJong et al. (2005) differs from that being currently used. Hence, in the model comparisons that follow, the DeJong et al. (2005) model was re-estimated using the new data set. Parameter estimates for the re-estimated model are available upon request.

The parameter estimates (conditional ML) for γ and ν presented in Table 4.2 indicate a strong error-correction effect. The autoregressive coefficient γ (0.3486 with std.err. 0.0391) indicates non-trivial persistence for shocks. This implies a fairly strong proportional error correction with an estimated ECM coefficient of 0.652 $(1-\gamma)$. The second ECM coefficient ν is estimated to be 0.2569 (s.e. 0.0483), indicating that past errors in the lagged tension index also affect the GDP growth noticeably. These results are generally in line with those obtained by DeJong et al. (2005), with coefficients (γ and ν) obtained from the current model being slightly higher that those obtained from DeJong et al. (2005). Estimates characterizing the distribution of $\ln a_j$, $\ln b_j$, $\overline{m}_j$, $\ln \sigma_j^2$ are also presented in Table 4.2. Note that there exists strong (and statistically significant) correlation between $\ln a_j$ and $\ln b_j$. Correlations among the remaining parameters are smaller. During model estimation, it was observed that the correlation between $\ln a_j$ and $\ln b_j$ is substantially higher in the relatively short regimes. The

mean and variance of volatility parameter $\ln\sigma_j^2$ were specified as being uncorrelated to the parameters governing the latent drift process. The estimated mean of $\ln\sigma_j^2$ is 2.0039. However, the variance of this parameter estimated to be 1.0142 has relatively high asymptotic standard error.

The smoothed means of the latent regime-parameters are illustrated in Figure 4.3. Note that the last five values of $\ln\sigma_j^2$ are lower than its mean value (2.0039), suggesting a definite reduction in the volatility of GDP growth. However, it is difficult to explicitly account for this trend in our model specification with only 17 available observations for the volatility. We are left with the possibility of a specification error in our model. However, the increased "sensitivity" of the latent-drift process implies that our model adjusts to new information very quickly. While we may begin with a conservative forecast for the volatility of growth in a new regime, the latent-process can adapt quickly to new information and correct its initial forecast as required. A conservative initial-estimate for the volatility protects us from making extreme assertions regarding the volatility of growth. Also, the model leaves the door open for possible increases in the volatility of growth in the future.

The parameter estimates from unconditional estimation are also presented in Table 4.2. In general, the parameters are fairly close to those obtained from the conditional estimation. The two variables with significant differences (using standardized differences computed as the ratio of the difference in the parameter estimates to the standard error of the conditional estimates) are the covariance between $\ln a_j$ and $\ln b_j$, and the covariance between $\ln b_j$ and $\overline{m}_j$. The estimate of the former from the conditional model has significantly smaller standard error relative to the standard error of the corresponding estimate from the unconditional model. The latter ($Cov(\ln b_j, \overline{m}_j)$), however, is not statistically significant in either scenario. A comparison of the parameters characterizing the latent-drift parameters reveals a general reduction in the velocity and acceleration. This indicates a reduction in the curvature of the latent drift process.

As shown in Table 4.1, the probabilities associated with the "optimal" regime-shift dates obtained by the conditional model are fairly high for most dates. Of the 17 regimes-change dates, the associated probability of 10 were greater than 90%, 3 were between 80 and 90% and

only 1 had probability weight less than 50% (1984:I). In contrast to the DeJong et al. (2005) model, the current model assigns a 76% probability that a regime change occurred in the third quarter of 2003. The DeJong et al. (2005) model does not detect a regime-shift at this quarter with any significant probability. To understand this difference, the top-right panel in Figure 4.2 provides a comparison of the latent-drift processes and the ECM component of GDP growth. While the latent-drift processes have similar trajectories in the two long regimes between 1984:II and 1999:IV, the last three regimes highlight the advantage of increased curvature of the latent-drift in the DHLR model. The latent-drift in the DHLR model provides a better fit to the ECM component, especially in the two short regimes between 2000:I and 2003:III. In the short regime 2000:I-2003:III the latent-drift corresponding to the DHLR model begins with a substantially higher initial velocity (as seen in the neighborhood of 2001:I) directly contributing its superior fit and its detection of a regime-shift at 2003:III. In contrast, the latent-drift process from DeJong et al. (2005) model is not sufficiently flexible to permit non-zero initial velocities for its latent-drift process.

4.3.2 Evolution of Regime-Shift Probabilities

It is important to note that the set of optimal regime-shift dates in Table 4.1 were obtained using the entire data set. In order to evaluate the ability of our model to anticipate future regime-shift dates, we present the evolution of probabilities associated with there being a break at the end-of-sample as new observations become available. For example, using the regime-shift dates up to 1984:I and growth observations up to 1989:IV, the conditional model was estimated. Probability of there being a regime-shift at each of the dates between 1984:II and 1989:IV (and there being none since 1984:I) was computed using the estimated parameters. Including the next observation (1990:IV) for growth, the above exercise was repeated to obtain the next set of regime-shift probabilities since 1984:I. Results from continuing this process until 1992:IV, illustrates the ability of the model to anticipate the regime-shift at 1991:I (see top-left panel of Figure 4.4). The entire exercise was repeated to predict each of the three regime-shifts after 1991:I (1999:IV, 2001:III and 2003:III).

The first panel in Figure 4.4 shows that the model predicts the regime-shift at 1991:I very rapidly. At the very quarter it was realized, it received a probability weight of nearly 75%. By the next quarter, the probability of no regime-shift decreases to a small number. Correspondingly, Figure 4.1 shows that the tension index turned very sharply around 1991:I. However, the detection of the regime-shift at 1999:IV was not quite as rapid. The probability of a regime-shift at 1999:IV (top-right panel of Figure 4.4) reaches a peak of nearly 85% about seven quarters after that date. Despite the slow increase of the regime-shift probability, alternative scenarios receive little weight. The regime-shift at 1999:IV also corresponds to a long and slow upward evolution of the tension index as seen in Figure 4.1 with the local peak not as pronounced as the local trough near 1991:I. The regime-shift at 2001:III received a maximal probability of nearly 70% five quarters after the event. With all three of the above regime-shift dates, we could have been fairly certain of there being a shift very rapidly as few alternative dates received much weight.

Prediction probabilities for the regime-shifts were re-computed using DeJong et al. (2005) model. Since a majority of the results do not change substantially from those reported in the DeJong et al. (2005) paper, we do not report the new results (except for the regime-shift at 2003:III). Figure 4.2 illustrates the behavior of the drift process in the DHLR model is similar to that in the DeJong et al. (2005) model during long regimes. In short regimes, the drift process in the DHLR model exhibits increased non-linearities. A direct implication of the increased non-linearities is apparent when we compare the ability of the two models in anticipating the regime-shift at 2003:III. Figure 4.5 illustrates the prediction probabilities of the regime-shift at 2003:III using the DeJong et al. (2005) model. Compared to the lower-right panel on Figure 4.4, it is immediately apparent that the DeJong et al. (2005) model does not assign any substantial weight to a regime-shift at 2003:III. Even with the GDP data for 2008:I, the probability accorded to a regime-shift at 2003:III is just a $\frac{1}{2}\%$. In contrast, the DHLR model (lower-right panel, Figure 4.4) assigns a peak probability of nearly 80% by 2006:IV. Note, from Figure 4.1, that the local peak in the tension index near 2003:III is substantially muted when compared to other peaks. However, the improved fit of the non-linear m_t process

(Figure 4.2) permits the model to detect a regime-shift at 2003:III.

4.3.3 Comparison of Forecasting Performance

In addition to detecting shifts of the economy from regimes of general expansion to general contraction (and vice-versa), we utilize the DHLR model to obtain forecasts of GDP growth. In light of the interesting debate in literature regarding the benefit conferred by non-linear models vis-à-vis forecasting, we present comparisons of forecasting performance of the DHLR model with the linearized version of the DHLR model, the DeJong et al. (2005) model and the random-walk model.

Along similar lines as Marcellino (2008), we obtain one-period-ahead out-of-sample forecasts of growth rate beginning from the first quarter of 1988 to the last quarter of 2007 (78 quarters). Beginning with the data set truncated at 1988:I, the DHLR model was estimated and a one-period-ahead forecast for GDP growth was obtained by aggregating over 50000 simulated draws of growth. With the inclusion of a new observation for the growth, the model was re-estimated and forecasts obtained from the re-estimated model. A similar procedure was used for obtaining forecasts from the DeJong et al. (2005) model and the random-walk models.

The linearized version of the DHLR model was created by replacing the non-linear m_t process with a linear approximation. Thus, for each of the terminal dates, the ECM component of growth $(y_t + \nu G_{t-2} - \gamma y_{t-2})$ was computed using the parameters (ν, γ) obtained from the DHLR model. To this, we fit a linear trend using ordinary least squares. The modified DGP corresponds to an ARMA(3,1) process with a linear drift. As above, forecasts were obtained by aggregating over 50000 simulated draws of one-period-ahead growth.

Forecasts from both the DeJong et al. (2005) and the linearized DHLR models are computed conditional on the sequence of regime-shift dates obtained by the DHLR model; making the forecast comparisons much more robust. Following conventional practice, we include the drifting random-walk model in our comparison. The long horizon of 78 quarters comprises of both long and short regimes. By 1988:I, the first of the two long regimes observed in the 1980s

and 1990s was well underway. During this and the next regimes, the DHLR and DeJong et al. (2005) models performed similarly when anticipating regime-shifts and this similarity is also observed in their forecasting performances.

Table 4.3 reports the root-mean-squared-error (RMSE) and root-variance-squared-error (RVSE) statistics of the forecast errors for each model. The RMSE values were computed as the average of the squared differences between the point forecasts of growth and the actual growth rates. RVSE was computed as the variance of the squared forecast errors. Since the forecast period extended across regimes, the regime-wise comparisons are presented first and the overall RMSE and RVSE can be found in the last two rows of the table. The DHLR and DeJong et al. (2005) models perform similarly during the long regimes. The difference in RMSE values between the two models is larger during the short regimes. The DHLR model outperforms the random-walk specification in all the regimes and in the overall comparison.

Importantly, the DHLR model has lower forecast errors relative to the linearized version; highlighting the added benefit of non-linearities in forecasting growth. We can also note that the variance of the squared forecast errors of the DHLR model are substantially smaller than those of the linearized version. Interestingly, the DeJong et al. (2005) model also outperforms the linearized DHLR model, reiterating the importance of non-linearities in capturing the trajectory of growth in different regimes.

4.3.4 Forecasting Future Growth Rates and Recessions

We present forecasts of future growth (conditional on a regime-shift having occurred at specific quarters) from both the DHLR and DeJong et al. (2005) models. Over an eight-quarter horizon beginning from 2008:II, forecasts are obtained by simulating trajectories of growth rate and regime-shift probabilities. For example, as the simulation of a sequence proceeds from the start of the forecast horizon, regime-shift occurs at a quarter with the probability computed using (4.2). Conditional on a regime-shift having occurred at a quarter, the latent-drift process is re-initialized using new draws of the latent variables; and the simulation continues.

Half a million trajectories of growth were simulated and were classified on the basis of the

occurrence of a regime-shift. Figure 4.6 illustrates the probabilities of a regime-shift occurring during the eight quarters beginning with 2008:II using the DHLR model. This probability represents the fraction of sequences that experienced a regime-shift at the specific quarter. The probability of a regime-shift reaches a peak at the third quarter of 2009. The probability of no regime-shift occurring during the eight-quarter horizon is just 2%. The lower panel in Figure 4.6 illustrates the trajectory of mean-growth, conditional on regime-shifts. Conditional on a regime-shift having occurred, the growth rate shifts to an upward trajectory. Also, later regime-shifts tend to produce further decreases of the growth rate into negative territory before the shift.

In contrast, the DeJong et al. (2005) model (Figure 4.7) predicts that a regime-shift is highly unlikely (85% probability). Compared to the DHLR model, the conditional growth trajectories generally show less severe negative growth rates before the shift and, correspondingly, lower positive growth rates after a regime-shift. This difference in forecasting behavior can be directly attributed to the behavior of the latent-drift process. As Figure 4.2 indicates, the latent drift process of the DHLR model "dips" much lower than that of the DeJong et al. (2005) model in the neighborhood of 2008. This increased curvature of the drift process leads to the realization of more negative growth rates, leading to a decrease in the tension index resulting in the higher possibility of a regime-shift.

Finally, we obtain forecasts of the probability of a recession in the near future by applying the deterministic recession dating procedure to forecast trajectories of growth and tension index. For each simulated growth trajectory, the optimal start-quarter for a recession was computed using the deterministic rule-based procedure. Conditional on a recession having started at the above quarter, the optimal end-quarter was also computed, automatically providing us the length of the regime.

Utilizing 500000 simulated trajectories of growth and the tension index (unconditional of regime-shifts), the probabilities of a recession starting at 2008:II is found to be nearly 44% by the DHLR model. As the top panel of Figure 4.8 indicates, the probability of the start of a recession does not fall below 10% until the second quarter of 2009. The probability that no

recession will occur is found to be around 15%. The lower panel in Figure 4.8 indicates that the recession would most likely be four quarters long (41%). Correspondingly, the middle panel shows most simulated recessions ending by 2010:III with the bulk of recession-end quarters clustered around 2009:III.

As noted above, the DeJong et al. (2005) model predicts lower probabilities of a regime shift and hence fewer trajectories with negative growth to cause a decrease in the tension index. Thus, only the trajectories with a later regime-shift lead to lower growth rates and decreases in the tension index. The top panel of Figure 4.9 shows that the DeJong et al. (2005) model typically predicts recessions starting much later with the most probable quarter being 2010:II. With typical recessions lasting 4-6 quarters (lower panel Figure 4.9), most recessions end in the neighborhood of 2011:II.

4.4 CONCLUSION

We have presented a non-linear regime-switching model of GDP growth that exhibits substantial non-linearities required for the accurate detection and prediction of transitions of U.S. economy from a recession to an expansion or vice versa. This increased non-linearity of the latent drift process governing growth in any given regime allows for a better fit of the model and substantially enhanced sensitivity to fluctuations in growth that might be indicative of regime-changes. The distributional characterization of the parameters governing the latent drift process is sufficiently broad to withstand mild trends in the latent parameters. Results from regime-shift prediction and in-sample forecasting exercises amply illustrates the benefits of the new model specification over the DeJong et al. (2005) model. We also present a deterministic rule-based method for dating recessions. Forecasts using this method indicates that a recession is highly likely in the next few quarters.

4.5 TABLES AND FIGURES

Table 4.1: Regime-Shift Dates and Estimated Probabilities

No.	Break Date	Estimated Probability	No.	Break Date	Estimated Probability
1	**1953:IV**	**0.991**	**10**	**1978:II**	**0.992**
2	**1955:I**	**0.949**	**11**	**1980:II**	**0.990**
3	**1958:I**	**0.998**		1983:IV	0.365
			12	**1984:I**	**0.432**
	1959:I	0.141		1984:II	0.103
4	**1959:II**	**0.859**			
			13	**1991:I**	**0.976**
5	**1960:IV**	**0.991**			
				1997:III	0.100
	1965.III	0.093		1998:IV	0.066
6	**1965:IV**	**0.680**	**14**	**1999:IV**	**0.638**
7	**1970:IV**	**0.952**		2001:II	0.035
			15	**2001:III**	**0.845**
8	**1973:I**	**0.898**		2001:IV	0.061
9	**1974.III**	**0.853**	**16**	**2003:III**	**0.766**
				2003:IV	0.075
				2004:III	0.073

Note: Regime-Shift Dates are in Boldface, other dates that receive non-negligible probability weights are also listed.

Table 4.2: Parameter Estimates

	Conditional ML		Unconditional ML	
Parameter	Estimate	Asym. Std. Error	Estimate	Asym. Std. Error
ν	-0.2569	0.0483	-0.2489	0.1168
γ	0.3486	0.0391	0.3239	0.0629
β_0	10.7660	0.8336	10.1935	1.1484
β_1	1.2396	0.0830	1.2894	0.2285
δ	0.5869	0.1792	0.5784	0.2639
$\ln a_j$	-2.7822	0.2390	-2.9235	0.6969
$\ln b_j$	-1.1173	0.2201	-1.3093	0.4154
$\overline{m}_j$	0.4116	0.1456	0.3098	0.1315
$\ln \sigma_j^2$	2.0039	0.2730	2.0150	0.5161
$Cov(\ln a_j, \ln b_j)$	0.6262	0.0390	0.5797	0.2020
$Cov(\ln a_j, \overline{m}_j)$	0.0042	0.0167	0.0173	0.01639
$Cov(\ln b_j, \overline{m}_j)$	-0.0314	0.0380	-0.0799	0.1858
$Var(\ln a_j)$	0.9045	0.2589	0.9287	0.3636
$Var(\ln b_j)$	0.4593	0.1599	0.4409	0.1629
$Var(\overline{m}_j)$	3.4960	0.4654	3.3574	0.5273
$Var(\ln \sigma_j^2)$	1.0142	0.5387	0.9717	0.5767
Log-Likelihood	-611.35554		-604.903793	

Note: The sequence of regime-shift dates used for Conditional Maximum Likelihood are reported in Table 4.1.

Table 4.3: In-Sample Forecast Performance.

	DHLR	DLR (2005)	m_t-Linear	Random Walk
	1988:II to 1991:I			
Root Mean Squared Error	2.169	2.175	2.653	2.476
Root Variance Squared Error	4.538	5.778	6.861	4.965
	1991:II to 1999:IV			
Root Mean Squared Error	1.637	1.718	2.065	2.196
Root Variance Squared Error	2.866	2.964	5.700	5.388
	2000:I to 2001:III			
Root Mean Squared Error	2.947	3.301	3.945	4.372
Root Variance Squared Error	7.186	9.938	8.350	15.992
	2001:IV to 2003:III			
Root Mean Squared Error	2.155	2.234	2.878	2.196
Root Variance Squared Error	8.065	8.613	10.617	5.380
	2003:IV to 2007:IV			
Root Mean Squared Error	1.525	1.628	2.265	2.221
Root Variance Squared Error	2.130	2.189	6.212	6.437
	1988:II to 2007:IV			
Root Mean Squared Error	1.904	2.013	2.503	2.509
Root Variance Squared Error	4.692	5.587	7.578	8.183

Forecast errors were computed using one-period-ahead forecasts of growth.

Figure 4.1: Tension Index and NBER-Recessions

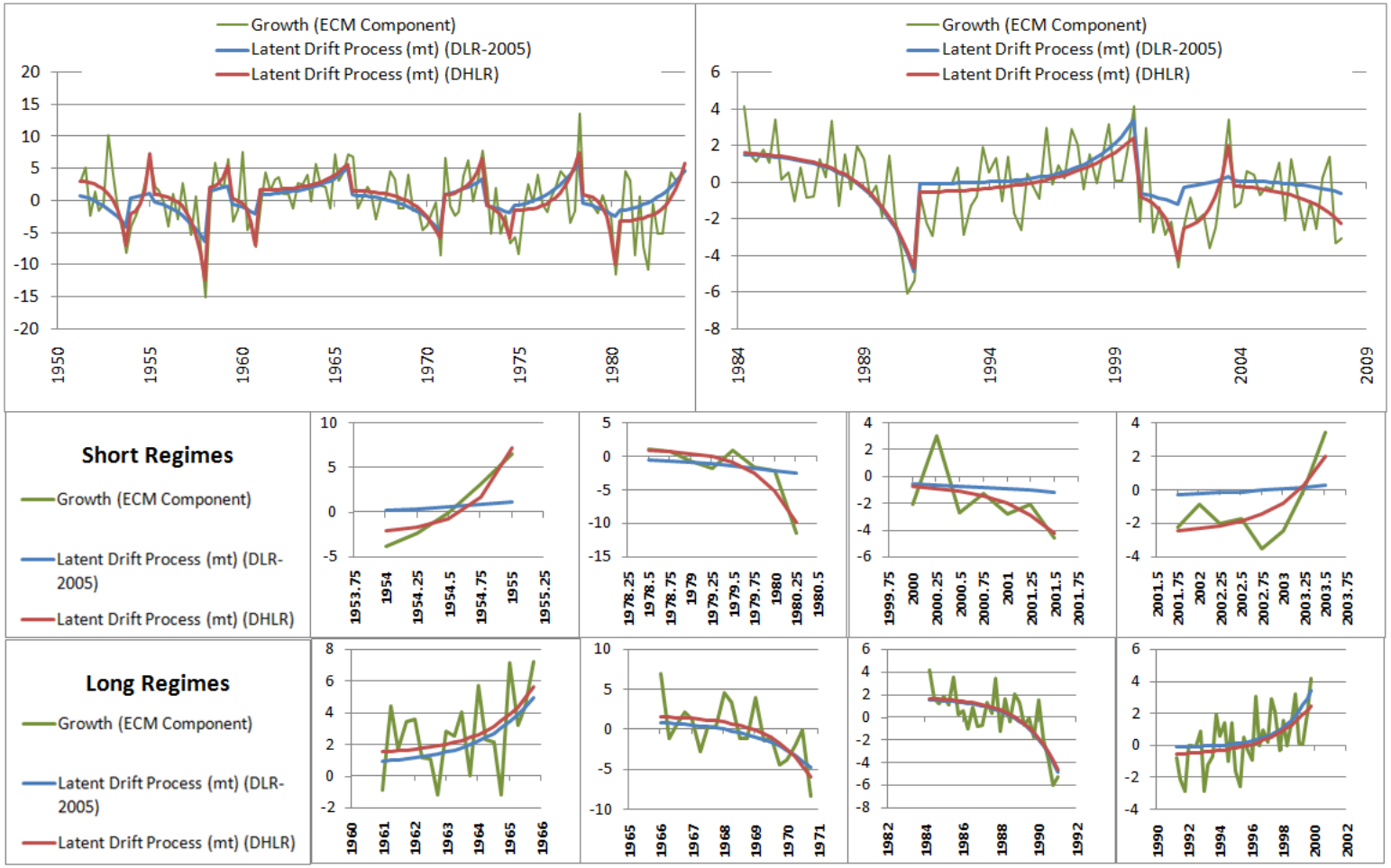


Figure 4.2: Comparison of the Latent Drift Process m_t.

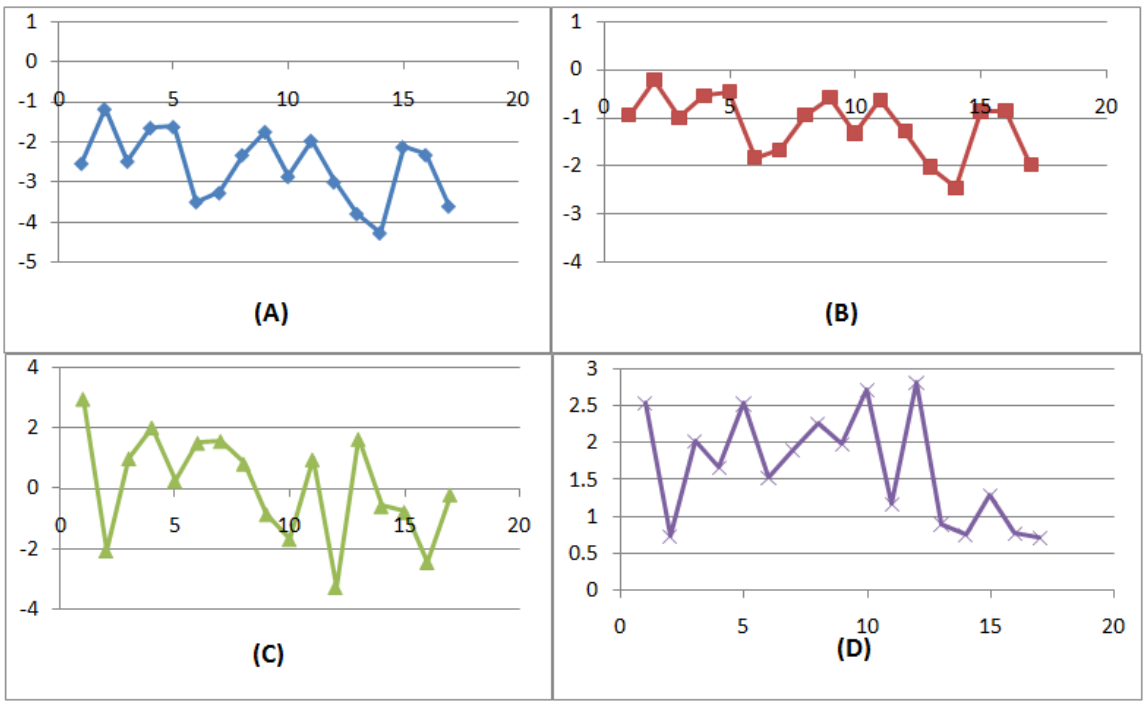


Figure 4.3: Smoothed Values of the Latent Variables.

Panel (A) $\rightarrow a_j$, **(B)** $\rightarrow b_j$, **(C)** $\rightarrow \overline{m}_j$, **(D)** $\rightarrow \ln \sigma_j^2$.

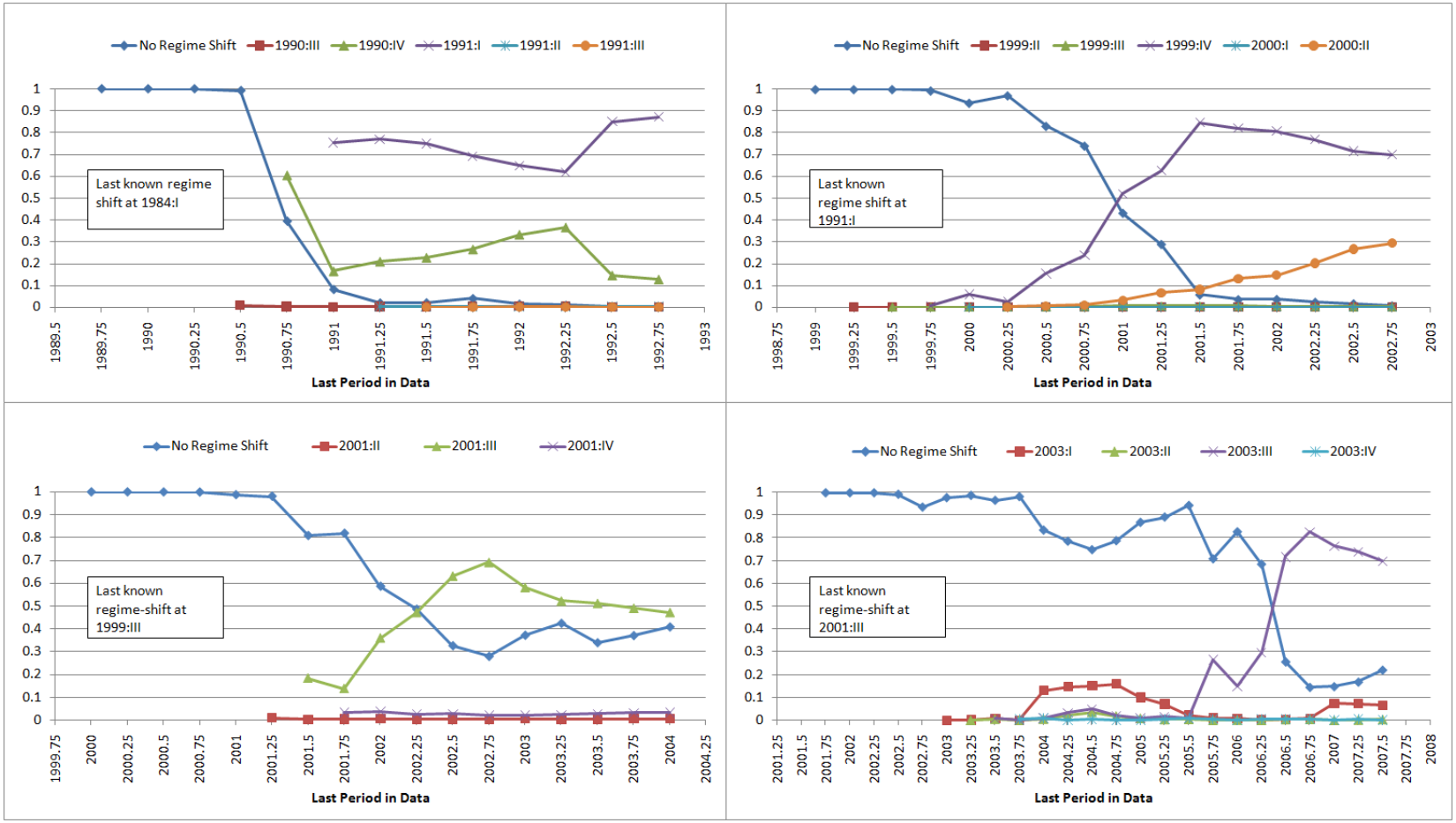


Figure 4.4: Predicting regime shifts using the DHLR Model.

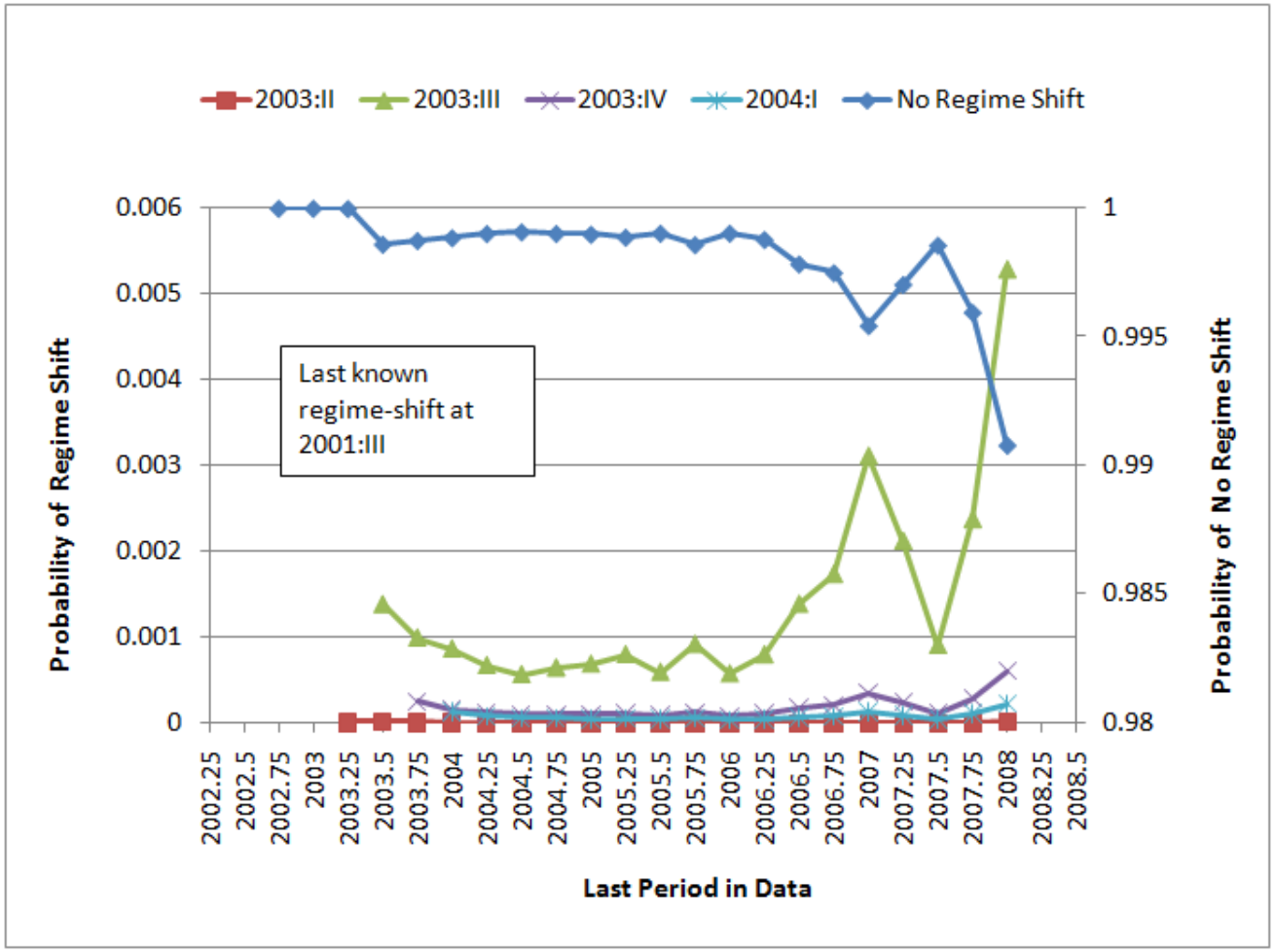


Figure 4.5: Predicting the regime shift at 2003:III using the DeJong et al. (2005) Model.

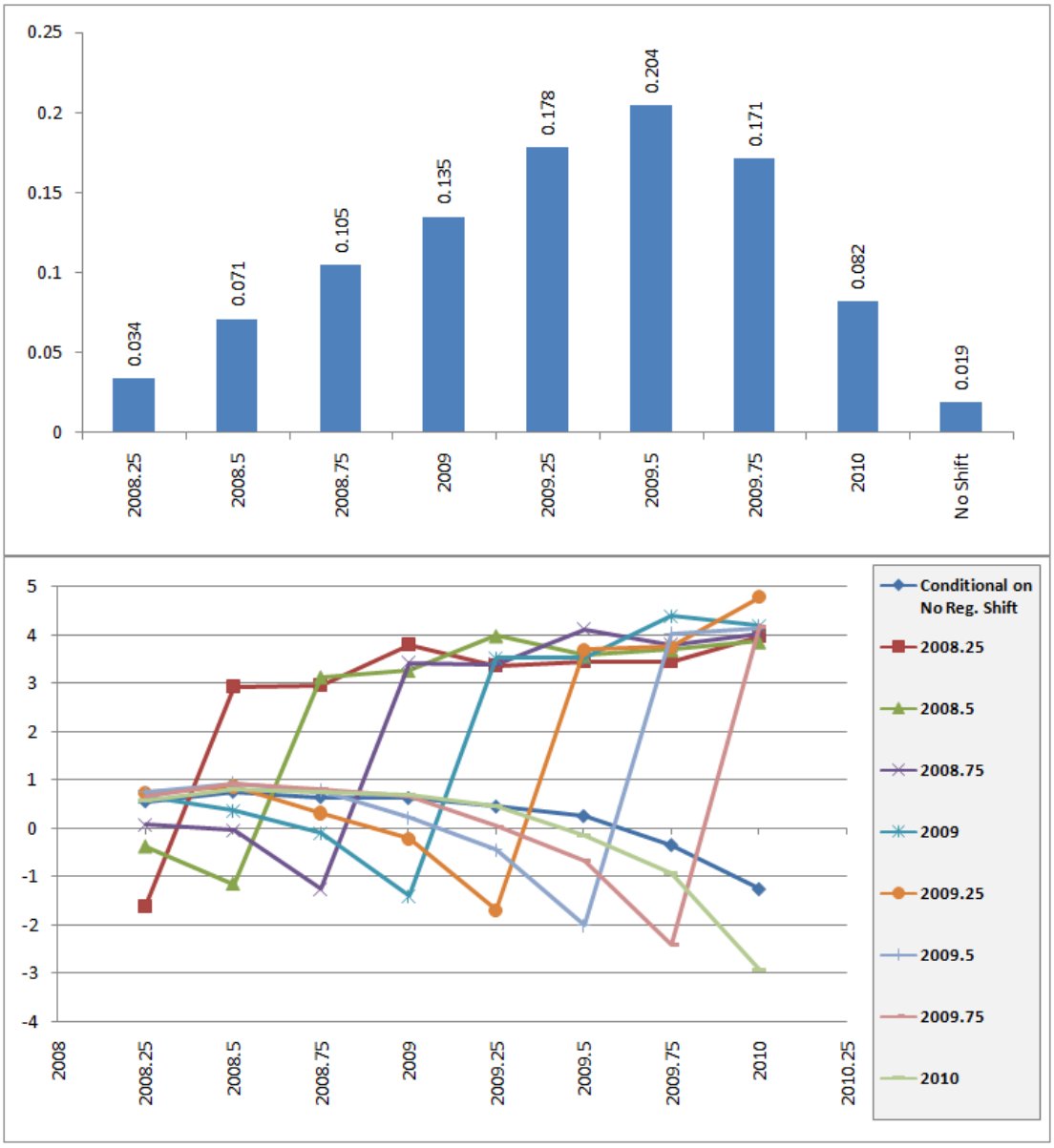


Figure 4.6: Forecasts of regime-shift probabilities and growth from the DHLR model.

Top Panel: Probability of a regime-shift at each quarter.
Bottom Panel: Mean forecast growth trajectories conditional on a regime shift at each quarter over the forecast horizon.

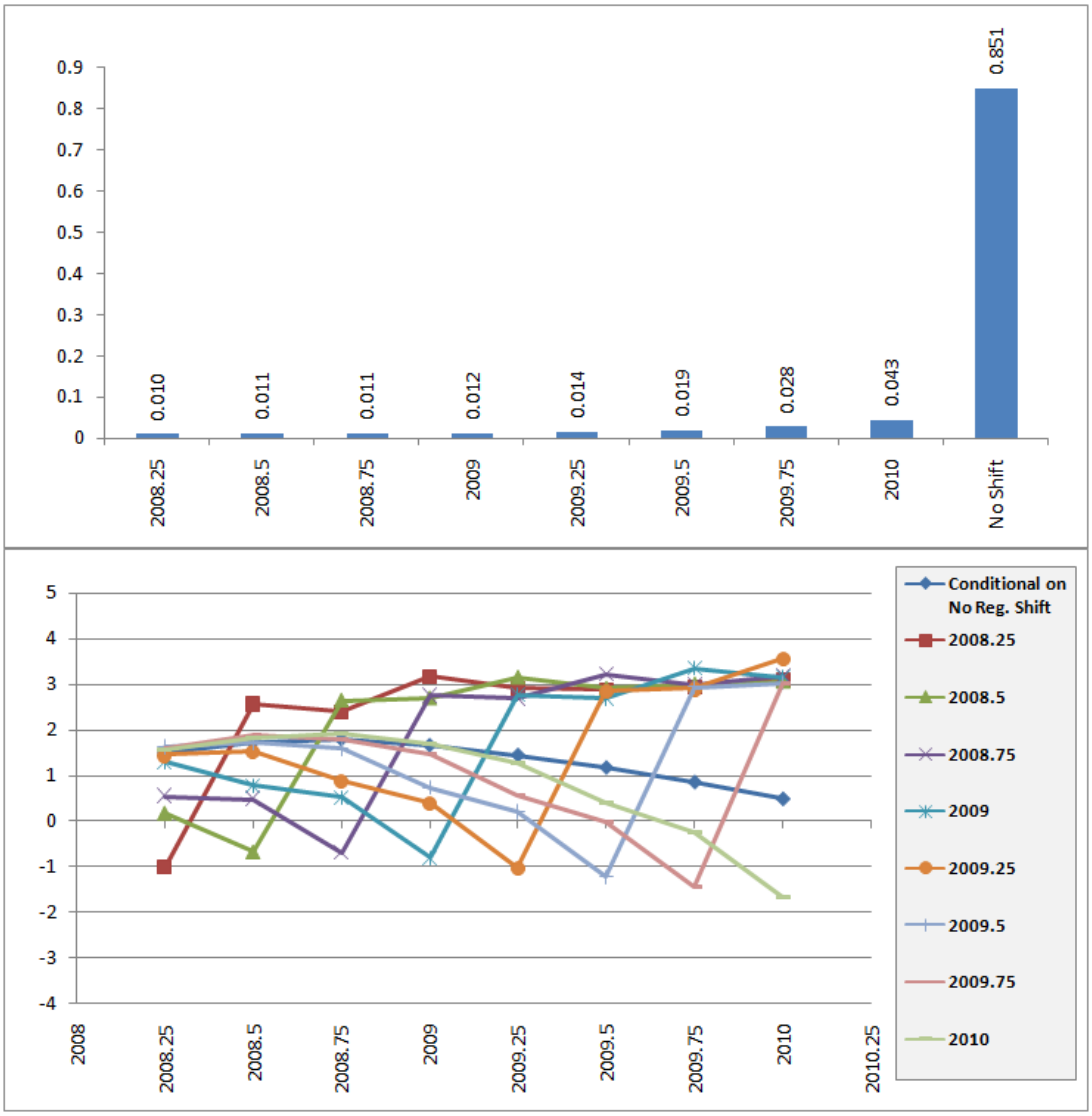


Figure 4.7: Forecasts of regime-shift probabilities and growth from the DeJong et al. (2005) model.

Top Panel: Probability of a regime-shift at each quarter.
Bottom Panel: Mean forecast growth trajectories conditional on a regime shift at each quarter over the forecast horizon.

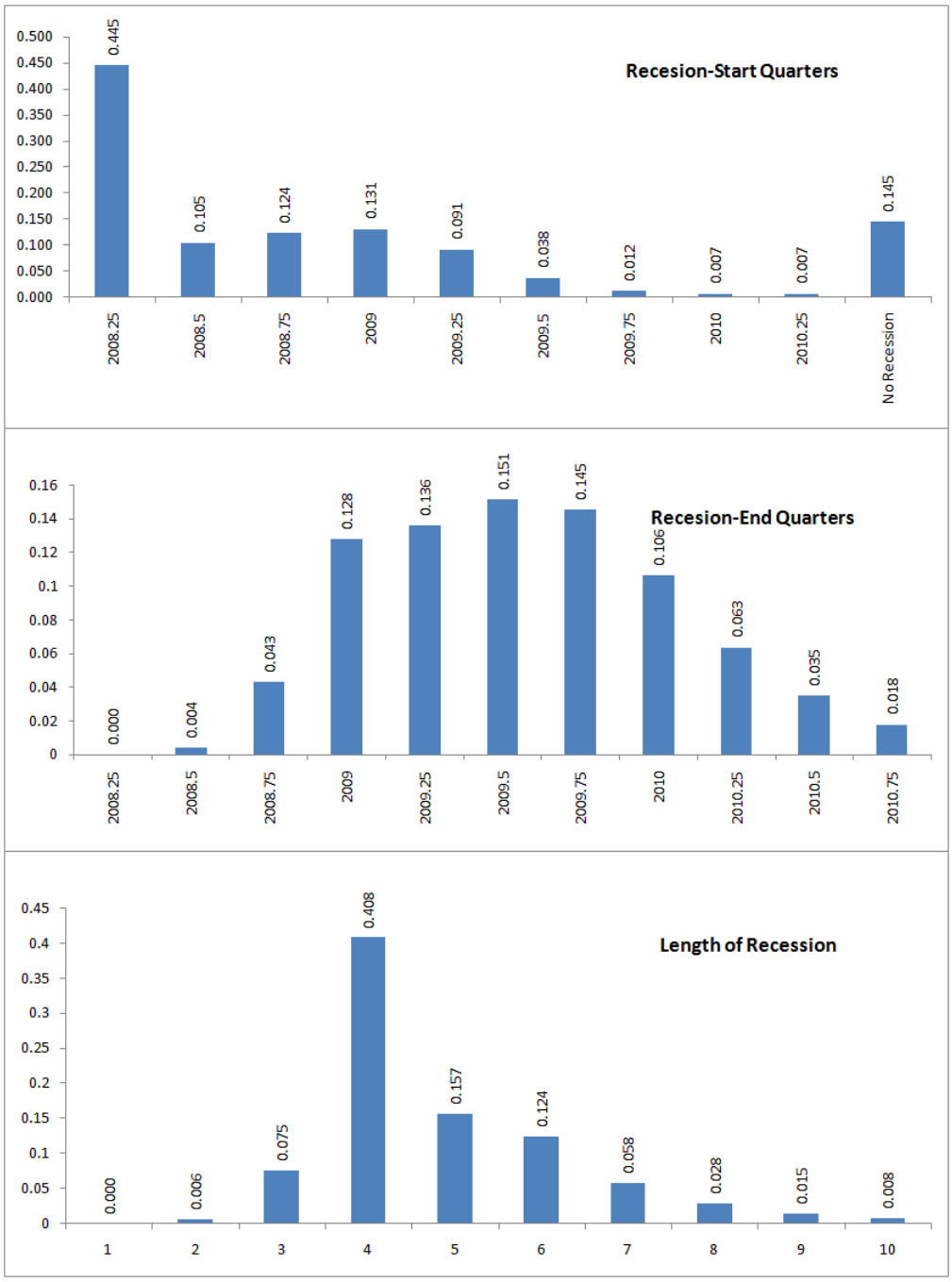


Figure 4.8: Recession Dating using DHLR Model.

Top Panel: Probability of a recession beginning at each quarter (including the probability of no recession in the forecast horizon).
Middle Panel: Conditional on a recession having started, the probability of recession ending at a particular quarter.
Bottom Panel: Probabilities of recession lengths.

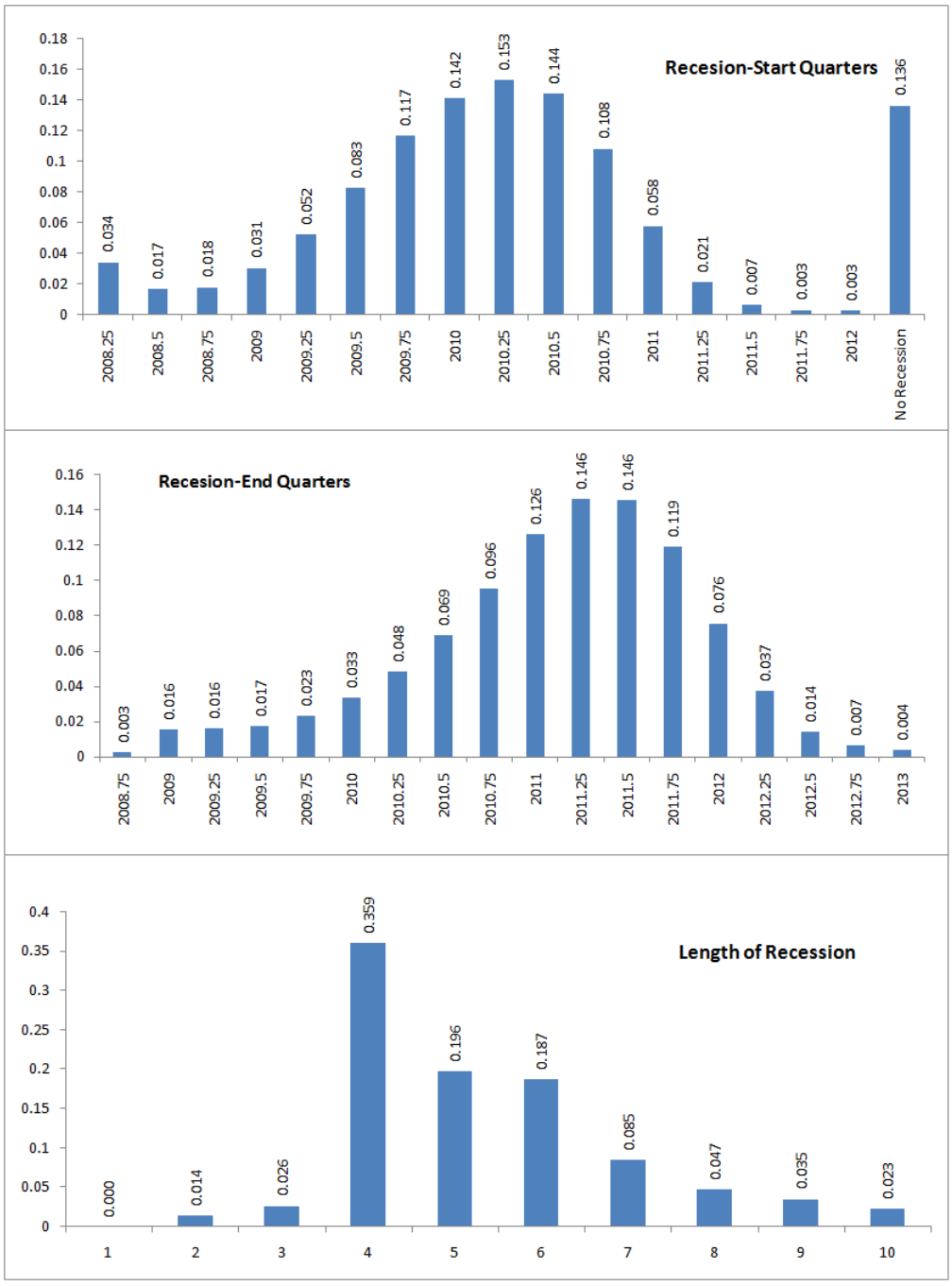


Figure 4.9: Recession Dating using DeJong et al. (2005) Model.

Top Panel: Probability of a recession beginning at each quarter (including the probability of no recession in the forecast horizon).
Middle Panel: Conditional on a recession having started, the probability of recession ending at a particular quarter.
Bottom Panel: Probabilities of recession lengths.

APPENDIX A

BEARINGS-ONLY TRACKING MODEL

A.1 SINGULAR CASE

Derivation of $\chi_t(\theta_t)$

For ease of notation we suppress t subscripts. The kernel $g_s(s)$ defined in (2.69) depends upon the quadratic form

$$\gamma(s) = \begin{pmatrix} \rho e_\theta \\ \beta \end{pmatrix}' P \begin{pmatrix} \rho e_\theta \\ \beta \end{pmatrix} - 2 \begin{pmatrix} \rho e_\theta \\ \beta \end{pmatrix} q. \tag{A.1.1}$$

We partition P and q conformably with $(\rho e'_\theta \quad \beta')$ into

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \qquad q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}. \tag{A.1.2}$$

Standard Gaussian algebra operations (square completion in β and ρ successively) produce the following expressions for $\gamma(s)$:

$$\gamma(s) = (\beta - b_\theta)' P_{22} (\beta - b_\theta) + a_\theta (\rho - r_\theta)^2 - s_\theta^2, \tag{A.1.3}$$

$$b_\theta = P_{22}^{-1} (q_2 - \rho P_{21} e_\theta), \qquad a_\theta = e'_\theta P_{11.2} e_\theta, \tag{A.1.4}$$

$$P_{11.2} = P_{11} - P_{12} P_{22}^{-1} P_{21}, \tag{A.1.5}$$

$$r_\theta = \frac{1}{a_\theta} \left(q_1 - P_{12} P_{22}^{-1} q_2\right)' e_\theta, \qquad s_\theta^2 = a_\theta r_\theta^2 + q'_2 P_{22}^{-1} q_2. \tag{A.1.6}$$

It follows that $\chi(\theta)$, as defined in (2.71), is given by

$$\chi(\theta) = 2\pi|P_{22}|^{-1} d_\theta \exp\left(\frac{1}{2}s_\theta^2\right), \tag{A.1.7}$$

$$d_\theta = \int_0^\infty \rho \exp\left(-\frac{1}{2}a_\theta(\rho - r_\theta)^2\right) d\rho. \tag{A.1.8}$$

Introducing the transformation of variables

$$\phi = \sqrt{a_\theta}(\rho - r_\theta), \tag{A.1.9}$$

d_θ can be written as

$$d_\theta = \frac{1}{a_\theta}\int_{-c_\theta}^\infty (\phi + c_\theta)\exp\left(-\frac{1}{2}\phi^2\right) d\phi \tag{A.1.10}$$

$$= \frac{1}{a_\theta}\left[\exp\left(-\frac{1}{2}c_\theta^2\right) + c_\theta\sqrt{\frac{\pi}{2}}\left(1 + \mathbf{erf}(\frac{c_\theta}{\sqrt{2}})\right)\right], \tag{A.1.11}$$

with $c_\theta = r_\theta\sqrt{a_\theta} > 0$, and $\mathbf{erf}()$ denoting the error function

$$\mathbf{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z \exp(-\phi^2) d\phi. \tag{A.1.12}$$

(The properties of $\mathbf{erf}()$ are discussed, e.g., in Abramowitz and Stegun, 1972). In deriving (A.1.11), we have exploited the fact that $r_\theta > 0$.

CRN-EIS draws of (β, ρ, θ)

An EIS draw of (β, ρ, θ) obtains from a CRN draw (u_1, u_2, u_3, u_4), where (u_1, u_2) denotes two $U(0,1)$ draws and (u_3, u_4) two i.i.d.N(0,1) draws, through the following sequence of transformations: (i) θ obtains from u_1 by inversion of the cdf associated with the piecewise loglinear EIS sampler $m(\theta)$. (ii) $\rho|\theta$ obtains from u_2 by inversion of the cdf associated with

$$m(\rho|\theta) = \frac{1}{d_\theta}\rho\exp\left(-\frac{1}{2}a_\theta(\rho - r_\theta)^2\right), \qquad \rho > 0. \tag{A.1.13}$$

Details of this transformation are provided below. (iii) $\beta|\rho, \theta$ obtains from the transformation

$$\beta = b_\theta + L\begin{pmatrix} u_3 \\ u_4 \end{pmatrix}, \tag{A.1.14}$$

where L denotes the Cholesky decomposition of P_{22}^{-1}.

Regarding step (ii), $\rho|\theta$ obtains from the transformation of (A.1.9), rewritten as

$$\rho = \frac{1}{\sqrt{a_\theta}}(\phi + c_\theta), \tag{A.1.15}$$

where the density of $\phi|\theta$ is given by

$$f_\phi(\phi|\theta) = \frac{1}{d_\theta a_\theta}(\phi + c_\theta)\exp\left(-\frac{1}{2}\phi^2\right), \qquad \phi > -c_\theta, \tag{A.1.16}$$

with cdf

$$\begin{array}{c} F_\phi(\phi|\theta) = \frac{1}{d_\theta a_\theta}\{\left[\exp\left(-\frac{1}{2}c_\theta^2\right) - \exp\left(-\frac{1}{2}\phi^2\right)\right] \\ +c_\theta\sqrt{\frac{\pi}{2}}\left[\mathbf{erf}\left(\frac{\phi}{\sqrt{2}}\right) + \mathbf{erf}\left(\frac{c_\theta}{\sqrt{2}}\right)\right]\} \end{array}, \tag{A.1.17}$$

accounting for the fact that $\mathbf{erf}(-z) = -\mathbf{erf}(z)$. For the application described in Section 5.4, c_θ turns out to be significantly larger than zero, so that ϕ is nearly $N(0,1)$. Thus for the inversion of the CRN $u_2 \sim U(0,1)$, we take as a starting value the corresponding (inverse) Gaussian draw $\phi^{(0)} \sim N(0,1)$ and iterate once or twice by Newton

$$\phi^{(k+1)} = \phi^{(k)} - \frac{F\left(\phi^{(k)}|\theta\right) - u_2}{F\left(\phi^{(k)}|\theta\right)}. \tag{A.1.18}$$

Derivation of $f(\lambda_{t+1}|Y_t)$

We again suppress t subscripts for ease of notation; accordingly, the index $t+1$ is replaced by $+1$. The product $g_\lambda(\lambda) f(\alpha_{+1}|\lambda)$ in (2.81) depends on the quadratic form

$$\delta(\alpha_{+1}, \lambda) = (\lambda' P\lambda - 2\lambda' q) + (\alpha_{+1} - A\lambda)'\Omega^{-1}(\alpha_{+1} - A\lambda). \tag{A.1.19}$$

It is transformed into a quadratic form in (α, λ_{+1}) via the inverse Dirac transformation (2.80)

$$\delta_1(\alpha, \lambda_{+1}) = \delta(\alpha_{+1}, \lambda)|_{\beta=\psi(\lambda_{+1},\alpha)}. \tag{A.1.20}$$

This implies the following two transformations:

$$\lambda|_{\beta=\psi(\lambda_{+1},\alpha)} = C\begin{pmatrix}\alpha\\ \lambda_{+1}\end{pmatrix}, \qquad (\alpha_{+1} - A\lambda)|_{\beta=\psi(\lambda_{+1},\alpha)} = D\begin{pmatrix}\alpha\\ \lambda_{+1}\end{pmatrix}, \tag{A.1.21}$$

(A.1.22)

with C and D respectively being 4×6 and 2×6 matrices partitioned in 2×2 blocks:

$$C=\begin{pmatrix} I_2 & 0 & 0 \\ -2I_2 & 2I_2 & -I_2 \end{pmatrix}, \qquad D=(I_2 \quad -I_2 \quad I_2).$$

Thus

$$\delta_1(\alpha,\lambda_{+1}) = \begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix}' M \begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix} - 2\begin{pmatrix} \alpha \\ \lambda_{+1} \end{pmatrix}' m, \tag{A.1.23}$$

$$M = C'PC + D'\Omega^{-1}D, \quad m = C'q. \tag{A.1.24}$$

Note that $\delta_1(\alpha,\lambda_{+1})$ is functionally similar to $\gamma(s)$ in (A.1.1), with β replaced by λ_{+1}. Therefore, the subsequent transformations of $\delta_1(\alpha,\lambda_{+1})$ are similar to those of $\gamma(s)$ outlined above, except that integration in (ρ,θ) is conditional on λ_{+1}, since it is $f(\lambda_{+1}|Y)$ that is now being evaluated. M and m are partitioned conformably with $\begin{pmatrix}\alpha' & \lambda'_{+1}\end{pmatrix}$ as

$$M=\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \qquad m=\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}.$$

After transformation from α to (ρ,θ), $\delta_1(\alpha,\lambda_{+1})$ becomes

$$\delta_*(\rho,\theta_{+1},\lambda_{+1}) = \lambda'_{+1}M_{22}\lambda_{+1} - 2\lambda'_{+1}m_2 + a^*_\theta(\rho - r^*_{\theta\lambda})^2 - c^*_{\theta\lambda}, \tag{A.1.25}$$

with

$$a^*_\theta = e'_\theta M_{11} e_\theta, \qquad r^*_{\theta\lambda} = \frac{1}{a^*_\theta}(m_1 - M_{12}\lambda_{+1})' e_\theta, \qquad c^*_{\theta\lambda} = r^*_{\theta\lambda}\sqrt{a^*_\theta}.$$

Regrouping (A.1.22) and integrating with respect to ρ, we obtain

$$\begin{aligned} f(\lambda_{+1}|Y) &= \frac{|\Omega|^{-\frac{1}{2}}}{2\pi}\exp\left\{-\frac{1}{2}(\lambda'_{+1}M_{22}\lambda_{+1} - 2\lambda'_{+1}m_2)\right\} \\ &\quad \times \int \left[\frac{1}{\chi(\theta)}d^*_{\theta\lambda}\exp\left(\frac{1}{2}c^{*2}_{\theta\lambda}\right)\right] m(\theta)\,d\theta, \end{aligned} \tag{A.1.26}$$

where $d^*_{\theta\lambda}$ obtains from (A.1.11) by substituting $\left(a^*_\theta, c^*_{\theta\lambda}\right)$ for (a_θ, c_θ). Since $m(\theta)$ is typically a tight density in our application, the variance of the terms between brackets under the integral sign is expected to be minimal, and the integral in (A.1.26) can be estimated accurately by MC using the same EIS draws from $m(\theta)$ used for the evaluation of ℓ_t.

A.2 NON-SINGULAR CASE

The computation of $\chi(\theta)$ and CRN-EIS draws of (β, ρ, θ) are the same as for the singular case. The derivation of $f(\lambda_{t+D}|Y_t)$ under the non-singular transition defined in (2.90) is straightforward. As above, we suppress the index t, and replace $t+D$ by $+D$. The product $g_\lambda(\lambda) f(\lambda_{+D}|\lambda)$ in (2.92) depends on the quadratic form

$$\delta(\lambda_{+D}, \lambda) = (\lambda' P \lambda - 2\lambda' q) + (\lambda_{+D} - A_D \lambda) V_D^{-1} (\lambda_{+D} - A_D \lambda), \tag{A.2.1}$$

which is rewritten as

$$\delta(\lambda_{+D}, \lambda) = (\lambda' P_0 \lambda - 2\lambda' q_0) + \lambda'_{+D} V_D^{-1} \lambda_{+D}, \tag{A.2.2}$$

with

$$P_0 = P + A'_D V_D^{-1} A_D, \qquad q_0 = q + A'_D V_D^{-1} \lambda_{+D}.$$

The integration with respect to λ in (A.1.26) proceeds exactly as described in the singular case, except that (P, q) are replaced by (P_0, q_0). Thus $f(\lambda_{+D}|Y)$ is given by

$$\begin{aligned} f(\lambda_{+D}|Y) &= \frac{|V_D|^{-\frac{1}{2}}}{2\pi} \exp\left\{ -\frac{1}{2} \left(\lambda'_{+D} V_D^{-1} \lambda_{+D} \right) \right\} \\ &\quad \times \int \left[\frac{1}{\chi(\theta)} d^0_{\theta\lambda} \exp\left(\frac{1}{2} \left(s^0_{\theta\lambda} \right)^2 \right) \right] m(\theta)\, d\theta, \end{aligned} \tag{A.2.3}$$

where $s^0_{\theta\lambda}$ and $d^0_{\theta\lambda}$ are defined by (A.1.6) and (A.1.8), with (P, q) replaced by (P_0, q_0). The EIS evaluation of (A.2.3) parallels that of $f(\lambda_+|Y)$ in (A.1.26).

APPENDIX B

PSEUDO-CODE FOR GAUSSIAN EIS

Without loss of generality, let the integral being computed be denoted as

$$\aleph = \int \varphi(x)dx. \tag{B.1.1}$$

With Gaussian-EIS, the importance sampler is a Gaussian distribution, i.e., $g(x;a)$ in equation (3.13) is a Gaussian density. The mean and variance of the sampler become the auxiliary parameters.

Step 1: (Initialize Sampler $g(x;a)$**)** Choose the initial value of the auxiliary parameters a^l, $l = 0$: the mean and variance of the gaussian density $g(x;a)$.

Step 2: (Recursive Optimization) The objective is to obtain optimal values of the auxiliary parameters. The following steps are repeated until convergence.

1. Draw R values of x from $g(x;a^l)$; denote these draws as $\left\{x_i^l\right\}_{i=1}^R$.
2. Obtain updated values of a^{l+1} as the solution to the least squares problem,

$$(a,c_o)^{l+1} = \arg\min_{a,c} \sum_{i=1}^{R} \left[\ln\varphi(x_i^l) - c_o - \ln g(x_i^l;a)\right]^2, \tag{B.1.2}$$

   where c_o is an intercept meant to calibrate $\ln\left(\frac{\varphi(x)}{g(x;a)}\right)$. (Details on the least-squares problem are provided below.)
3. Check for convergence.

Once convergence is reached, we have the optimal mean and variance of the EIS sampling density: $\widehat{a}$. As emphasized by DeJong et al. (2007), draws of x from the importance sampler are obtained by transforming a fixed set of random numbers from a canonical distribution (in the present case, standard normal). The set of canonical draws are referred to as "Common Random Numbers" or CRNs.

Step 3: (Likelihood Evaluation) Draw N values $\{x_i\}_{i=1}^{N}$ from the optimal EIS sampling density $g(x;\widehat{a})$. The IS estimate of the integral and the EIS weights are given by

$$\widehat{\aleph}_N = \frac{1}{N}\sum_{i=1}^{N}\omega_t^i, \tag{B.1.3}$$

$$\omega_i = \frac{\varphi(x_i)}{g(x_i;\widehat{a})}. \tag{B.1.4}$$

Details on the Auxiliary Regression in Gaussian-EIS

Let x be a j-dimensional variable with elements $(x_1, x_2, ...x_j)$. Then the auxiliary parameters a are the $j \times 1$ vector of means and the $j \times j$ covariance matrix. Since the covariance matrix is symmetric, the number of auxiliary parameters reduces to $j + j(j+1)/2$.

We take a^l as given, initialized by a^0. Hereafter, we will drop the superscript l and describe a single iteration of the auxiliary regression.

Let the mean vector associated with a be denoted by μ, and the precision matrix (the inverse of the covariance matrix) by H. The setup of the LS problem arises from the approximation $\ln\varphi(x)$ by a gaussian kernel:

$$\begin{aligned}\ln\varphi(x) &\propto -\frac{1}{2}(x-\mu)'H(x-\mu)\\ &\propto -\frac{1}{2}\left(x'Hx - 2x'H\mu\right).\end{aligned}$$

The term $x'Hx$ can be written as,

$$\begin{pmatrix} x_1 & x_2 & . & . & x_j \end{pmatrix} \begin{pmatrix} h_{11} & h_{21} & . & . & h_{1j} \\ h_{21} & h_{22} & . & . & h_{2j} \\ . & . & . & . & . \\ . & . & . & . & . \\ h_{j1} & h_{j2} & . & . & h_{jj} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_j \end{pmatrix}$$

$$
\begin{aligned}
= \quad & h_{11}\left(x_1^2\right) + h_{22}\left(x_2^2\right) + \ldots + h_{jj}\left(x_j^2\right) \\
& + 2h_{21}\left(x_2 x_1\right) + 2h_{31}\left(x_3 x_1\right) + \ldots + 2h_{j1}\left(x_j x_1\right) \\
& + 2h_{32}\left(x_3 x_2\right) + 2h_{42}\left(x_4 x_2\right) + \ldots + 2h_{j2}\left(x_j x_2\right) \\
& \ldots \\
& + 2h_{j(j-1)}\left(x_j x_{j-1}\right).
\end{aligned}
$$

From the above decomposition it is evident that the coefficients of the squares, pairwise products and the individual components of x are in one-to-one correspondence with the means and precision matrix of the gaussian approximation. Thus, the LS problem reduces to the regression of $\ln \varphi(x)$ on

$\left[1, x_1^2, x_2^2, \ldots, x_j^2,\ x_1 x_2, x_1 x_3, \ldots, x_{j-1} x_j,\ x_1, \ldots, x_j\right]$. For a j-dimensional variable x, the number of regressors is $\left(1 + j + \frac{j(j+1)}{2}\right)$.

As a concrete example, consider a 3-dimensional problem where $x = [x_1, x_2, x_3]$. The regression reduces to

$$
\begin{aligned}
\ln \varphi(s) \quad = \quad & \lambda_0 + \lambda_1\left(x_1^2\right) + \lambda_2\left(x_2^2\right) + \lambda_3\left(x_3^2\right) \\
& + \lambda_4\left(x_2 x_1\right) + \lambda_5\left(x_3 x_1\right) + \lambda_6\left(x_3 x_2\right) \\
& + \lambda_7 x_1 + \lambda_8 x_2 + \lambda_9 x_3.
\end{aligned}
$$

Having obtained the LS estimates, the updated precision matrix is given by

$$
\begin{aligned}
h_{11} \quad = \quad & -2\lambda_1;\ h_{22} = -2\lambda_2;\ h_{33} = -2\lambda_3 \\
h_{21} \quad = \quad & -\lambda_4;\ h_{31} = -\lambda_5;\ h_{32} = -\lambda_6.
\end{aligned}
$$

The updated means can be obtained by using the coefficients $(\lambda_7, \lambda_8, \lambda_9)$:

$$
\mu = H^{-1} \begin{pmatrix} \lambda_7 \\ \lambda_8 \\ \lambda_9 \end{pmatrix}.
$$

When s_t is univariate, the LS problem reduces to the regression

$$\ln \varphi(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2.$$

The updated mean and variance can be written as

$$\begin{aligned} \sigma^2 &= -\frac{1}{2\lambda_2}, \\ \mu &= -\frac{1}{2}\frac{\lambda_1}{\lambda_2}. \end{aligned}$$

APPENDIX C

EIS WITH THE META-GAUSSIAN SAMPLER (EXAMPLE 1)

As above, we denote the integral to be computed as

$$\aleph = \int \varphi(x)dx. \tag{C.1.1}$$

The importance sampler ($g(x;a)$ in equation 3.13) consists of two piecewise-continuous density approximations for the margins and the meta-Gaussian copula. The following sequence of operations describes the process of computing the optimal meta-Gaussian importance sampler.

Step 1: (Initialize Sampler $g(x;a)$)

1. The margins are initialized by choosing R uniformly spaced nodes in the space of each margin over which the piecewise density approximation will be constructed.
2. The copula parameter (ρ) is initialized (typically to zero). R draws of $\widetilde{x}$ $(= [\widetilde{x}_1, \widetilde{x}_2]')$ are obtained from a bivariate normal density with zero mean and correlation ρ by the transformation of a fixed set of random numbers (CRNs).
3. Draws of $\widetilde{x}_1$ and $\widetilde{x}_2$ are converted to x_1 and x_2 using the transformation in equation (3.19). At this stage, the marginal distributions are simply uniform over the chosen intervals.

Step 2: (Piecewise approximations for the margins) The procedure for creating the piecewise-continuous density approximation is described in DeJong et al. (2007).

1. The first step requires the computation of the target density at the R nodes in the uniformly spaced grid in each margin. This is accomplished using kernel density estimation.
2. Subsequent steps replicate the process outlined in DeJong et al. (2007).

Having created the piecewise density approximations for the margins, they remain fixed and are not updated in the ensuing steps.

Step 3: (The optimal correlation ρ) With the distributions of the margins determined, the optimal correlation parameter is computed by a process similar to the Gaussian-EIS outlined above. Given any value of the correlation parameter ρ, draws from the copula density are obtained by the following procedure.

1. Draw R values of $\widetilde{x}$ from a bivariate normal density with zero mean and correlation ρ by the transformation of a fixed set of random numbers (CRNs).
2. Convert draws of $\widetilde{x}$ to x using the inverse of the transformation in equation (3.19).

Analogous to the procedure outlined above, the setup of the LS problem arises from the approximation $\ln \varphi(x)$ by the bivariate meta-Gaussian density (3.21). Omitting the parameters associated with the marginals

$$\ln \varphi(x) - \ln(f_1(x_1)) - \ln(f_2(x_2)) \propto \frac{-1}{2(1-\rho^2)} \left[\widetilde{x}_1^2 - 2\rho \widetilde{x}_1 \widetilde{x}_2 + \widetilde{x}_2^2\right] \tag{C.1.2}$$

Practical experience suggested that the omission of the marginal densities did not adversely affect the estimation of ρ. Hence, the least squares problem simplifies to the regression

$$\ln \varphi(x) = \lambda_0 + \lambda_1 \widetilde{x}_1 \widetilde{x}_2 \tag{C.1.3}$$

Updated values of ρ can be obtained by transforming the LS coefficient λ_1. From the above expression

$$\lambda_1 = \frac{\rho}{1-\rho^2}. \tag{C.1.4}$$

The resulting quadratic equation in ρ has two real roots.

$$\rho = \frac{\sqrt{4\lambda_1^2 + 1} \pm 1}{2\lambda_1} \tag{C.1.5}$$

It can be noted that one root is the reciprocal of the other. Hence, one of them is always between -1 and 1. This process is repeated until convergence in ρ.

Step 4: (Likelihood Evaluation) Draw N values $\{x_i\}_{i=1}^{N}$ from the optimal meta-Gaussian copula sampler using the procedure outlined above. The IS estimate of the integral and the EIS weights are given by

$$\widehat{I_N} = \frac{1}{N}\sum_{i=1}^{N}\omega_t^i, \tag{C.1.6}$$

$$\omega_i = \frac{\varphi(x_i)}{g(x_i;\widehat{a})} \tag{C.1.7}$$

where $g(x;\widehat{a})$ is the meta-Gaussian importance sampling density.

APPENDIX D

INITIALIZING THE EIS PROCEDURE

The measurement equations for $\breve{y}_t$ and $\breve{i}_t$ can be linearized around their steady state values as

$$\begin{bmatrix} \breve{y}_t \\ \breve{i}_t \end{bmatrix} = \begin{bmatrix} \psi_y^0 \\ \psi_i^0 \end{bmatrix} + \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \end{bmatrix} \begin{bmatrix} \breve{K}_t \\ \ln \breve{A}_t \\ \ln \breve{V}_t \end{bmatrix} + \begin{bmatrix} \eta_y \\ \eta_i \end{bmatrix}. \tag{D.1.1}$$

Components of the matrix M are obtained using the derivatives of 3.37 and 3.38 with respect to the state variables evaluated at steady state:

$$M_{11} = \left[\frac{\partial}{\partial \breve{K}_t}\left(\breve{A}_t(u_t\breve{K}_t)^{\alpha}\right)\right]_{ss} = \frac{\alpha\bar{\breve{y}}}{\bar{\breve{K}}} \tag{D.1.2}$$

$$M_{12} = \left[\frac{\partial}{\partial \breve{A}_t}\left(\breve{A}_t(u_t\breve{K}_t)^{\alpha}\right)\right]_{ss}\left[\frac{\partial \breve{A}_t}{\partial \ln \breve{A}_t}\right]_{ss} = \bar{\breve{y}} \tag{D.1.3}$$

$$M_{13} = \frac{\partial \breve{A}_t(u_t\breve{K}_t)^{\alpha}}{\partial \ln \breve{V}_t} = 0 \tag{D.1.4}$$

$$M_{21} = \left[\frac{\partial \breve{i}_t}{\partial \breve{K}_t}\right]_{ss} = M_{11} - \left[\frac{\partial \breve{c}_t}{\partial \breve{K}_t}\right]_{ss} \tag{D.1.5}$$

$$M_{22} = \left[\frac{\partial \breve{i}_t}{\partial \ln \breve{A}_t}\right]_{ss} = M_{12} - \left[\frac{\partial \breve{c}_t}{\partial \breve{A}_t}\right]_{ss}\bar{\breve{A}} \tag{D.1.6}$$

$$M_{23} = \left[\frac{\partial \breve{i}_t}{\partial \ln \breve{V}_t}\right]_{ss} = \left[\frac{\partial \breve{c}_t}{\partial \breve{V}_t}\right]_{ss}\bar{\breve{V}}. \tag{D.1.7}$$

Hence, the linearized versions of the measurement equations can be written as

$$\breve{y}_t = \psi_y^0 + M_{11}\breve{K}_t + M_{12}\breve{A}_t + \eta_y \tag{D.1.8}$$

$$\breve{i}_t = \psi_i^0 + M_{21}\breve{K}_t + M_{22}\breve{A}_t + M_{22}\breve{V}_t + \eta_i, \tag{D.1.9}$$

where, the constants ψ_y^0 and ψ_i^0 are given by

$$\psi_y^0 = \breve{y}(1-\alpha-\ln\overline{\breve{A}}) \tag{D.1.10}$$

$$\psi_i^0 = \breve{i} - M_{21}\overline{\breve{K}} - M_{22}\ln\overline{\breve{A}} - M_{23}\ln\overline{\breve{V}}. \tag{D.1.11}$$

The derivatives of the consumption function with respect to latent variables can be computed by finite difference of the Chebyshev polynomial approximation created earlier. The linearized measurement equations provide an approximate conditional density for the observable variables. A density for the state variables conditional on the current observations can be obtained by combining the approximate (Gaussian) measurement density obtained above.

The linear approximation to the measurement equations can be written as

$$\begin{bmatrix} \breve{y}_t - \psi_y^0 \\ \breve{i}_t - \psi_i^0 \end{bmatrix} \sim N_2\left(\Psi s_t, \Xi\right), \tag{D.1.12}$$

where

$$\Psi = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \end{bmatrix}, \tag{D.1.13}$$

and

$$\Xi = \begin{pmatrix} \sigma_y^2 & 0 \\ 0 & \sigma_i^2 \end{pmatrix}. \tag{D.1.14}$$

Let the approximate density for the state variables be written as

$$f(s_t) \approx N_3\left(\mu, \Omega\right). \tag{D.1.15}$$

The joint density can then be written as

$$f\begin{pmatrix} \breve{y}_t - \psi_y^0 \\ \breve{i}_t - \psi_i^0 \\ s_t \end{pmatrix} \approx N_5\left(\begin{bmatrix} \Psi\mu \\ \mu \end{bmatrix}, \begin{bmatrix} \Xi + \Psi\Omega\Psi' & \Omega\Psi' \\ \Psi\Omega & \Omega \end{bmatrix}\right). \tag{D.1.16}$$

Transforming the above density provides a conditional (Gaussian) density for the state variables with a mean

$$\mu + \Psi\Omega\left[\Xi + \Psi\Omega\Psi'\right]^{-1}\left(\begin{bmatrix} \breve{y}_t - \psi_y^0 \\ \breve{i}_t - \psi_i^0 \end{bmatrix} - \Psi\mu\right) \tag{D.1.17}$$

and variance

$$\Omega - \Psi\Omega\left[\Xi + \Psi\Omega\Psi'\right]^{-1}\Omega\Psi'. \tag{D.1.18}$$

Generally, utilizing a Gaussian approximation for the period-$(t-1)$ filtering density ($f(s_{t-1}|X_{t-1})$, linearization of the state-transition equations provides a Gaussian approximation to the prediction density depicted in equation (D.1.15). In the DSGE example presented, $\ln\breve{A}_t$ and $\ln\breve{V}_t$ are independent of the past. Hence, corresponding Gaussian approximations are readily available. The transition in $\breve{K}_t$ alone requires linearization and can be accomplished by a procedure similar to that illustrated above.

BIBLIOGRAPHY

Abramowitz, M., and I. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Washington : U.S. Govt. Print. Off., 1972, 10th printing, 1972 edition.

An, S., and F. Schorfheide. "Bayesian Analysis of DSGE Models.", 2007. Econometric Reviews, Forthcoming.

Ang, A., G. Bekaert, and M. Wei. "Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better?" Working Paper 11538, National Bureau of Economic Research, 2005. http://www.nber.org/papers/w11538.

Arias, A., G. D. Hansen, and L. E. Ohanian. "Why Have Business Cycle Fluctuations Become Less Volatile?" Working Paper 12079, National Bureau of Economic Research, 2006. http://www.nber.org/papers/w12079.

Beaudry, P., and C. Koop. "Do recessions permanently change output?" *Journal of Monetary Economics* 31: (1993) 149–163.

Blanchard, O., and J Simon. "The Long and Large Decline in U.S. Output Volatility." *Brooking Papers on Economic Activity* 1: (2001) 135–164.

Bordo, M.D., and T. Helbling. "Have National Business Cycles Become More Synchronized?" Working Paper 10130, National Bureau of Economic Research, 2003. http://www.nber.org/papers/w10130.

Bouezmarni, T., and J.V.K. Rombouts. "Semiparametric Multivariate Density Estimation for Positive Data Using Copulas.", 2007. HEC Montreal Working Paper.

Bouyé, E., and M. Salmon. "Dynamic Copula Quantile Regressions and Tail Area Dynamic Dependence in Forex Markets.", 2002. http://www2.warwick.ac.uk/fac/soc/wbs/research/wfri/rsrchcentres/ferc/wrkingpaprseries/. University of Warwick, Working Paper.

Brannas, K., and J.G. De Gooijer. "Autoregressive-asymmetric moving average models for business cycle data." *Journal of Forecasting* 13: (1994) 529–544.

Brendstrup, B., and H. J. Paarsch. "Semiparametric Identification and Estimation in Multi-Object English Auctions." *Journal of Econometrics* 141: (2007) 84–108.

Brunner, A. D. "Conditional Asymmetries in Real GNP: A Seminonparametric Approach." *Journal of Business & Economic Statistics* 10: (1992) 65–72.

Bry, G., and C. Boschan. *Cyclical analysis of time series: selected procedures and computer programs*. NBER: New York., 1971.

Burns, A., and W. Mitchell. *Measuring Business Cycles*. New York: National Bureau of Economic Research, 1947.

Cameron, A. C., T. Li, P. K. Trivedi, and D. M. Zimmer. "Modeling the differences in counted outcomes using bivariate copula models: With application to mismeasured counts." *Econometrics Journal* 7: (2004) 566–584.

Cameron, A. C., P. K. Trivedi, F. Milne, and J. Piggott. "A microeconomic model of the demand for health care and health insurance in Australia." *Review of Economic Studies* 55: (1988) 85–106.

Carpenter, P. Clifford, J.R., and P. Fernhead. "An Improved Particle Filter for Non-Linear Problems." *IEE Proceedings-Radar, Sonar and Navigation* 146(1): (1999) 2–7.

Cecchetti, S. G., A. Flores-Lagunes, and S. Krause. "Assessing the Sources of Changes in the Volatility of Real Growth." Working Paper 11946, National Bureau of Economic Research, 2006. http://www.nber.org/papers/w11946.

Chatterjee, S. "Capital utilization, economic growth and convergence." *Journal of Economic Dynamics and Control* 29: (2005) 2093–2124.

Chauvet, M., and J. D. Hamilton. "Dating Business Cycle Turning Points." Working Paper 11422, National Bureau of Economic Research, 2005. http://www.nber.org/papers/w11422.

Chauvet, M., and J. Piger. "A Comparison of the Real-Time Performance of Business Cycle Dating Methods." *Journal of Business and Economic Statistics* 26(1): (2008) 42–49.

Cherubini, U., E. Luciano, and W. Vecchiato. *Copula Methods in Finance*. West Sussex:John Wiley & Sons, 2004.

Chib, S., and R. Winkelmann. "Markov chain Monte Carlo analysis of correlated count data." *Journal of Business and Economic Statistics* 19: (2001) 428–435.

Christiano, L. J., and J. M. Davis. "Two Flaws In Business Cycle Accounting." Working Paper 12647, National Bureau of Economic Research, 2006. http://www.nber.org/papers/w12647.

Davis, S. J., and J. A. Kahn. "Interpreting the Great Moderation: Changes in the Volatility of Economic Activity at the Macro and Micro Levels." Working Paper 14048, National Bureau of Economic Research, 2008. http://www.nber.org/papers/w14048.

DeGroot, M.H. *Probability and Statistics*. Reading PA: Addison-Wesley, 1984.

DeJong, D. N., R. Liesenfeld, and J.F. Richard. "A Non-Linear Forecasting Model of GDP Growth." *Review of Economics and Statistics* 87(4): (2005) 97–708.

DeJong, D.N., D. Hariharan, R. Liesenfeld, and J.F. Richard. "An Efficient Approach to Analyzing State-Space Representations.", 2007. University of Pittsburgh Working Paper.

DeJong, D.N., B.F. Ingram, and C.H. Whiteman. "A Bayesian Approach to Dynamic Macroeconomics." *Journal of Econometrics* 98: (2000) 203–233.

DeJong, with C. Dave, D.N. *Structural Macroeconometrics*. New Jersey:Princeton University Press, 2007.

Demarta, S., and A. J. McNeil. "The t Copula and Related Copulas." *International Statistical Review* 73(1): (2005) 111–129.

Devroye, L. *Non-Uniform Random Variate Generation*. New York: Springer, 1986.

Diebold, F.X., J.-H. Lee, and G.C. Weinbach. "Regime Switching with Time-Varying Transition Probabilities." In *Nonstationary Time Series Analysis and Cointegration*, edited by Hargraves C., Oxford University Press: Oxford, 1994, 283:302.

Diebold, F.X., and G. Rudebusch. "A nonparametric investigation of duration dependence in the American business cycle." *Journal of Political Economy* 98: (1990) 596–616.

Doucet, A., N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. New York: Springer, 2001.

Fang, H.B., K.T. Fang, and S. Kotz. "The Meta-elliptical Distributions with Given Marginals." *Journal of Multivariate Analysis* 82: (2002) 1–16.

Fernandez-Villaverde, J., and J.F. Rubio-Ramirez. "On the Solution of the Growth Model with Investment-Specific Technological Change'.", 2004a. University of Pennsylvania Working Paper.

———. "Sequential Monte Carlo Filtering: An Example.", 2004b. University of Pennsylvania Working Paper.

———. "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood." *Journal of Applied Econometrics* 20: (2005) 891–910.

———. "Estimating Macroeconomic Models: A Likelihood Approach.", 2007. Forthcoming in Review of Economic Studies.

Fernandez-Villaverde, J., J.F. Rubio-Ramirez, and M.S. Santos. "Convergence Properties of the Likelihood of Computed Dynamic Models." *Econometrica* 74: (2006) 93–119.

Filardo, A. "Business Cycle Phases and Their Transitional Dynamics." *Journal of Business & Economic Statistics* 12: (1994) 299–308.

Fisher, J. "Technology Shocks Matter.", 2003. Mimeo, Federal Reserve Bank of Chicago.

French, M. W., and D. E. Sichel. "Cyclical Patterns in the Variance of Economic Activity." *Journal of Business & Economic Statistics* 11: (1993) 113–119.

Galvao, A.B.C. "Can non-linear time series models generate US business cycle asymmetric shape?" *Economics Letters* 77: (2002) 187–194.

Geweke, J. "Bayesian Inference in Econometric Models Using Monte Carlo Integration." *Econometrica* 24: (1989) 1037–1399.

Ghysels, E., A.C. Harvey, and E. Renault. "Stochastic Volatility." In *Handbook of Statistics Vol. 14.*, edited by G. Maddala, and C.R. Rao, Amsterdam: Elsevier, 1996.

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

Gordon, R. J. "What Caused the Decline in U.S. Business Cycle Volatility?" Working Paper 11777, National Bureau of Economic Research, 2005. http://www.nber.org/papers/w11777.

Gradshteyn, I. S., and I. M.Ryzhik. *Table of integrals, series, and products*. New York: Academic Press, 1965, 4th edition edition.

Greenwood, J., Z. Hercowitz, and P. Krusell. "The role of inversment-specific technological change in the business cycle." *European Economic Review* 44: (2000) 91–115.

Gregoir, S., and F. Lenglart. "Measuring the probability of a business cycle turning point by using a multivariate qualitative hidden Markov model." *Journal of Forecasting* 19: (2000) 81–102.

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

———. *Time Series Analysis*. Princeton: Princeton University Press, 1994.

———. "What's Real About the Business Cycle?" Working Paper 11161, National Bureau of Economic Research, 2005. http://www.nber.org/papers/w11161.

Hamilton, J.D., and G. Perez-Quiros. "What do leading indicators lead." *Journal of Business* 69: (1996) 27–49.

Harding, D., and A. Pagan. "A Comparison of Two Business Cycle Dating Methods." *Journal of Economic Dynamics and Control* 27: (2002a) 1681–1690.

———. "Synchronization of Cycles." *Journal of Econometrics* 132: (2006) 59–79.

Harding, D., and A.R. Pagan. "Dissecting the cycle: A methodological investigation." *Journal of Monetary Economics* 49: (2002b) 365–381.

———. "A Suggested Framework for Classifying the Modes od Cycle Research." *Journal of Applied Econometrics* 20: (2005) 151–159.

Heinen, A., and E. Rengifo. "Multivariate Modelling of Time Series Count Data: An Autoregressive Conditional Poisson Model.", 2003. http://www.core.ucl.ac.be/services/COREdp03.html. CORE Discussion Paper 2003/25.

Hendry, D.F. "Monte Carlo Experimentation in Econometrics." In *The Handbook of Econometrics, Vol. IV*, edited by R.F. Engle, and D.L. McFadden, New York: North Holland, 1994.

Hess, G.D., and S. Iwata. "Asymmetric persistence in GDP? A deeper look at depth." *Journal of Monetary Economics* 40: (1997a) 535–554.

———. "Measuring and comparing business-cycle features." *Journal of Business and Economic Statistics* 15: (1997b) 432–444.

Hornstein, A., and P. Krusell. "Can Technology Improvements Cause Productivity Slowdowns?" *NBER Macroeconomics Annual* 11: (1996) 209–259.

Ireland, P. "A Method for Taking Models to Data." *Journal of Economic Dynamics and Control* 28: (2004) 1205–1226.

Jacquier, E., N.G. Polson, and P.E. Rossi. "Bayesian Analysis of Stochastic Volatility Models." *Journal of Business and Economic Statistics* 12: (1994) 371–417.

Jaimovich, N., and H. E. Siu. "The Young, the Old, and the Restless: Demographics and Business Cycle Volatility." Working Paper 14063, National Bureau of Economic Research, 2008. http://www.nber.org/papers/w14063.

Joe, H. *Multivariate Models and Dependence Concepts*. New York: Chapman & Hall, 1997.

Jones, M.C., J. S. Marron, and S. J. Sheather. "A Brief Survey of Bandwidth Selection for Density Estimation." *Journal of the American Statistical Association* 91 (433): (1996) 401–407. http://www.jstor.org/stable/2291420.

Judd, K. L. *Numerical Methods in Economics*. Cambridge: The MIT Press, 1998.

Keilis-Borok, V., J.H. J. H. Stock, A. Soloviev, and P. Mikhalev. "Pre-recession Pattern of Six Economic Indicators in the USA." *Journal of Forecasting* 19: (2000) 65–80.

Kelly, K. S., and R. Krzysztofowicz. "A bivariate meta-Gaussian density for use in hydrology." *Stochastic Hydrology and Hydraulics* 11: (1997) 17–31.

KIM, C.-J., J. Morley, and J. Piger. "Bayesian Counterfactual Analysis Of The Sources Of The Great Moderation." *Journal of Applied Econometrics* 23: (2008) 173–191.

Kim, C.-J., and C.R. Nelson. "Business Cycle Turning Points, a New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model With Regime Switching." *Review of Economics and Statistics* 80: (1998) 188–201.

———. *State-space Model With Regime Switching*. Cambridge:MIT Press, 1999a.

Kim, C.J., and C. R. Nelson. "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of Business Cycle." *Review of Economics and Statistics* 81: (1999b) 1–10.

Kim, G., M. J. Silvapulle, and P. Silvapulle. "Comparison of semiparametric and parametric methods for estimating copulas." *Computational Statistics and Data Analysis* 51: (2007) 2836–2850.

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

Kitagawa, G. "Non-Gaussian State-Space Modeling of Non-Stationary Time Series." *Journal of the American Statistical Association* 82: (1987) 1032–1041.

———. "Monte Carlo Filter and Smoother for Non-Gaussian Non-Linear State-Space Models." *Journal of Computational and Graphical Statistics* 5: (1996) 1–25.

———. "A Self-Organizing State-Space Model." *Journal of the American Statistical Association* 93 (443): (1998) 1203–1215.

Koskinen, L., and L.-E Oller. "A Classifying Procedure for Signalling Turning Points." *Journal of Forecasting* 23: (2004) 197–214.

Lahiri, K., and J.G. Wang. "Predicting cyclical turning points with a leading index in a Markov switching model." *Journal of Forecasting* 13: (1994) 245–263.

Liesenfeld, R., and J.F. Richard. "Univariate and multivariate stochastic volatility models: estimation and diagnostics." *Journal of Empirical Finance* 10: (2003) 505– 531.

———. "Improving MCMC Using Efficient Importance Sampling.", 2006. University of Pittsburgh Working Paper.

Long, D., and R. Krzysztofowicz. "A Family of Bivariate Densities constructed from Marginals." *Journal of the American Statistical Association* 90 No. 430: (1995) 739–746.

Marcellino, M. "A Linear Benchmark for Forecasting GDP Growth and Inflation?" *Journal of Forecasting* 27(4): (2008) 305–340.

Mari, D. D., and S. Kotz. *Correlation and Dependence*. London: Imperial College Press, 2004.

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

Miranda, M.J., and P.L. Fackler. *Applied Computational Economics and Finance*. Cambridge: The MIT Press, 2002.

Nelder, J.A., and R. Mead. "A Simplex Method for Function Minimization." *Computer Journal* 7: (1965) 308–313.

Nelsen, R. B. *An Introduction to Copulas*. New York: Springer, 2006.

Niemira, M.P., and P.A. Klein. *Forecasting Financial and Economic Cycles.* Wiley: New York, 1994.

Otrok, C. "On Measuring the Welfare Cost of Business Cycles." *Journal of Monetary Economics* 47: (2001) 61–92.

Pagan, A. "The Getting of Macroeconomic Wisdom.", 1999. Discussion Paper, Australian National University.

Park, B. U., and J. S. Marron. "Comparison of data-driven bandwidth selectors." *Journal of the Americal Statistical Association* 85: (1990) 66–72.

Patton, A. J. "Copula-Based Models for Financial Time Series.", 2007. http://www.economics.ox.ac.uk/members/andrew.patton/. Oxford-Man Institute of Quantitative Finance Working Paper (OMI11/07).

Pesaran, M. H., and S. M. Potter. "A Floor and Ceiling Model of US Output." *Journal of Economic Dynamics and Control* 21:4-5: (1997) 661–695.

Phillips, A.W. "Stabilization Policy in a Closed Economy." *Economic Journal* 64: (1954) 290–333.

Phillips, A.W. "Stabilization Policy and the Time Form of Lagged Response." *Economic Journal* 67: (1957) 265–277.

Pitt, M.K. "Smooth Particle Filters for Likelihood Evaluation and Maximisation.", 2002. University of Warwick Working Paper.

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

Potter, S. "A Nonlinear Approach to U.S. GNP." *Journal of Applied Econometrics* 10: (1995) 109–126.

Potter, S. M. "A Nonlinear Model of the Business Cycle." *Studies in Nonlinear Dynamics & Econometrics* 4(2): (2000) 85–93.

Press, W.H., S.A. Teukolsky, W. T. Vetterling, and B.P. Flannery. *Numerical Recipes in Fortran 90*. Cambridge University Press: Cambridge, 1996.

Richard, J.-F., and W. Zhang. "Efficient High-Dimensional Monte Carlo Importance Sampling." *Journal of Econometrics* 141(2): (2007) 1385–1411.

Sargent, T.J. "Two Models of Measurements and the Investment Accelerator." *Journal of Political Economy* 97: (1989) 251–287.

Schweizer, B., and A. Sklar. *Probabilistic Metric Spaces*. New York: North Holland, 1983.

Sheather, S. J. "Density Estimation." *Statistical Science* 19: (2004) 588–597. http://www.jstor.org/stable/4144429.

Sheather, S. J., and M. C. Jones. "A reliable data-based bandwidth selection method for kernel density estimation." *Journal of the Royal Statistical Society. Series B (Methodological)* 53(3): (1991) 683–690. http://www.jstor.org/stable/2345597.

Sklar, A. "Fonctions de ŕepartition à *n* dimensions et leurs marges." *Publ. Inst. Statist Univ. Paris* 8: (1959) 229–231.

———. "Random variables, joint distributions, and copulas." *Kybernetica* 9: (1973) 449–460.

Smith, J.Q., and A.A.F. Santos. "Second-Order Filter Distribution Approximations for Financial Time Series with Extreme Outliers." *Journal of Business and Economic Statistics* 24: (2006) 329–337.

Smith, M. "Modeling selectivity using archimedean copulas." *Econometrics Journal* 6: (2003) 99–123.

Stock, J. H., and M. W. Watson. "New indexes of leading and coincident economic indicators." *NBER Macroeconomics Annual* 351–394.

Stock, J.H., and M. W. Watson. "Understanding Changes in International Business Cycle Dynamics." Working Paper 9859, National Bureau of Economic Research, 2003. http://www.nber.org/papers/w9859.

———. "Why Has U.S. Inflation Become Harder to Forecast?" Working Paper 12324, National Bureau of Economic Research, 2006. http://www.nber.org/papers/w12324.

Stock, J.H., and M.W. Watson. "Has the Business Cycle Changed and Why?" *NBER Macroeconomics Annual* 17: (2002) 159–218.

Taylor, S. J. "Financial Returns Modeled by the Product of Two Stochastic Processes: A Study of Daily Sugar Prices." In *Time Series Analysis : Theory and Practice Vol. 1*, edited by O.D. Anderson, Amsterdam: North Holland, 1982.

Taylor, S.J. *Modeling Financial Time Series*. Chichester: John Wiley and Sons, 1986.

Tong, H. *Non-linear Time Series: A Dynamical System Approach.* Oxford: Oxford University Press, 1990.

Trivedi, P. K., and D. M. Zimmer. "Copula Modeling: An Introduction for Practitioners'." *Foundations and Trends in Econometrics* 1 (1): (2005) 1–111.

———. "Using Trivariate Copulas to Model Sample Selection and Treatment Effects: An Application to Family Health Care Demand." *Journal of the American Statistical Association* 24(1): (2006) 63–76.