

Efficient Importance Sampling in Applied Econometrics

Inaugural-Dissertation

zur Erlangung des akademischen Grades eines Doktors
der Wirtschafts- und Sozialwissenschaften
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Guilherme Valle Moura, MA
aus Belo Horizonte, Brasilien
Kiel, 2009

**Gedruckt mit Genehmigung der Wirtschafts-
und Sozialwissenschaftlichen Fakultät der
Christian-Albrechts-Universität zu Kiel**

Dekan: Prof. Dr. Thomas Lux
Erstberichterstattender: Prof. Dr. Roman Liesenfeld
Zweitberichterstattender: Prof. Dr. Thomas Lux
Tag der Abgabe der Arbeit: 22.10.2009
Tag der mündlichen Prüfung: 08.01.2010

Abstract

This thesis focusses on econometric applications requiring multivariate numerical integration. Models that attempt to capture real-world complexities are typically nonlinear and display many unobservable factors. These characteristics imply that the likelihood function of these models contain high-dimensional integrals that often cannot be solved analytically, and thus have to be approximated numerically. Importance Sampling is a Monte Carlo simulation method often used to solve high-dimensional integration. In the present work, the Efficient Importance Sampling method developed by Richard and Zhang (2007) was used to overcome the problem of finding good multivariate importance samplers for the integrand of likelihood functions from different models. It was shown how importance sampling can be used to efficiently solve high dimensional integration problems in econometric problems involving panel data, and time series.

Contents

1	Introduction	13
1.1	Outline	19
2	Numerical Integration	25
2.1	Deterministic Methods of Integration	26
2.1.1	Univariate Quadrature Methods	28
2.1.1.1	Newton-Cotes quadratures	29
2.1.1.2	Gaussian quadratures	30
2.1.2	Multivariate Quadrature Methods	33
2.2	Monte Carlo Integration	35
2.2.1	Classical Monte Carlo Integration	37
2.2.2	Importance Sampling	40
2.2.2.1	GHK importance sampler	45
2.2.3	Efficient Importance Sampling	49
2.2.3.1	EIS for the exponential family of distributions	52
2.2.3.2	Sequential EIS	55
3	Estimation of Dynamic Panel Probit Models	59
3.1	Introduction	60
3.2	Determinants and Dynamics of Current Account Reversals	62
3.2.1	Determinants	62
3.2.1.1	State dependence	63
3.2.1.2	Serially correlated error terms	64
3.2.2	Data	65
3.3	Empirical Specifications	66
3.3.1	Random country-specific effects	67

CONTENTS

3.3.2	Random country- and time-specific effects	68
3.4	Maximum-Likelihood Estimation	69
3.4.1	Random country-specific effects	70
3.4.2	AR(1) country-specific errors	71
3.4.2.1	Estimation via the GHK algorithm	72
3.4.2.2	Efficient Estimation via the EIS algorithm	73
3.4.3	AR(1) time-specific effects	77
3.5	Empirical Results	79
3.5.1	A note on normalization	79
3.5.2	Model 1: Pooled Probit	79
3.5.3	Model 2: Random country-specific effects	83
3.5.4	Model 3: AR(1) country-specific errors	85
3.5.5	Model 4: AR(1) time-specific effects	88
3.5.6	Predictive Performance	91
3.6	Conclusion	94
3.A1	Appendix 1: EIS-implementation	97
3.A2	Appendix 2: MC experiments on the numerical efficiency of EIS	100
4	Nonlinear State-Space Models	105
4.1	Likelihood Evaluation and Filtering in State-Space Representations	107
4.2	The Particle Filter and Leading Extensions	109
4.3	Parametric EIS Filter	112
4.3.1	EIS integration	113
4.3.2	Continuous approximations of $f(s_t Y_{t-1})$	115
4.3.3	Degenerate transitions	116
4.4	Application to DSGE Models	118
4.4.1	Example 1: Two-State RBC Model	118
4.4.2	Example 2: Six-State Small Open Economy Model	138
4.4.3	Repeated Samples	145
4.5	Conclusion	148
4.A	Appendix 1: Nonlinear Approximate Solution of DSGE Models	148
4.A-1	The Stochastic Growth Model	149
4.A-2	Solution Methods	151
4.A-2.1	Perturbation	152
4.A-2.2	Projection Methods	154

CONTENTS

4.A-2.3	Finite Elements Method	155
4.A-2.4	Spectral Method	157
4.A-3	Numerical Comparisons	159
5	Conclusions	165
	References	180

CONTENTS

List of Figures

3.1	Average ℓ -Step Ahead Marginal Effects	87
3.2	ROC Curve	92
3.3	Duration of Reversal Episodes	94
4.1	Sample Impoverishment	111
4.2	Log-Likelihood Approximations	126
4.3	Sample Impoverishment in Practice I	128
4.4	Likelihood Cuts	130
4.5	Date-by-Date Filter Comparisons	133
4.6	Linear vs. Nonlinear Measurement Densities	135
4.7	Sample Impoverishment in Practice II	137
4.A-1	Benchmark Calibration	161
4.A-2	Extreme Calibration	162

LIST OF FIGURES

List of Tables

3.1	ML estimates of Model 1: Pooled probit	82
3.2	ML estimates of Model 2: Random country-specific effects	84
3.3	ML-EIS estimates of Model 3: AR(1) country-specific errors.	86
3.4	Model 3: Comparison between GHK and EIS ML estimation	89
3.5	ML-EIS estimates of Model 4: AR(1) time-specific effects	90
3.6	Classification errors and predicted average duration in years	95
3.7	List of countries	102
3.8	List of countries	103
4.1	Parameter Values, RBC Model	124
4.2	Estimates of RBC Model using simulated data.	131
4.3	Parameter Values, SOE Model	143
4.4	Monte Carlo Means and Standard Deviations, SOE Model	144
4.5	EIS Filter Repeated Samples	146
4.6	Repeated VAR Samples	147
4.A-1	Parameter Values, Stochastic Growth Model	159
4.A-2	Empirical Moments, Benchmark Calibration	160
4.A-3	Empirical Moments, Extreme Calibration	163
4.A-4	Computational Time	163

LIST OF TABLES

Chapter 1

Introduction

Economic and econometric models that attempt to better capture the complexities inherent to real-world economic behavior often cannot be solved analytically using algebra and calculus.

Models that lack closed-form solution are not unique to economics, and since the introduction of digital computers, scientists from different fields have been taking advantage of numerical methods to approximate solutions to their analytically intractable models. The progress in computing capabilities has been generating, and will continue to generate, tremendous new opportunities to both economics and econometrics.

In the case of economics, the power of modern computing makes it possible to analyze and solve an increasingly large collection of far more complex and realistic models. Model economies enriched with uncertainty and dynamics typically give rise to functional equations in which the unknowns are not anymore a set of points in the Euclidean space, but rather a function defined on a continuum of points. This is the case in Dynamic Stochastic General Equilibrium (DSGE) models, where the Bellman and Euler equations characterizing the optimal path of the economy are functional equations, as well as the conditions describing rational expectation and arbitrage pricing market equilibria. Only in very limited special cases, these functional equations allow a closed-form solution, implying that most models have to be solved numerically. Blanchard and Kahn (1980) developed an algorithm that allows the solution of linear approximations to DSGE models. Nevertheless, many economic models are not suitable to linear approximations, demanding significant computational work. Given the recent advances, it is now

1. INTRODUCTION

possible to solve DSGE models using nonlinear methods based on sophisticated extensions of interpolation procedures (see Aruoba et al., 2006, or the Appendix to Chapter 4).

Standard DSGE models are based on strong assumptions about the agents' ability to analyze, collect, and process information in order to make forecasts about economic variables. The development of the digital computer allowed the use Monte Carlo (hereafter also referred to as MC) simulation methods to approximate the solution of intractable models. These methods have been used in economics to analyze complex economies populated by heterogeneous agents with different beliefs, and consequently different behaviors (see Axtell, 2000).

In econometrics, massive improvements in data storage capacity and access to technology have opened up new classes of data sets that could be analyzed (see Hamilton, 2006). These includes high-frequency data, from both tick-by-tick transactions in financial markets as well as from retail scanners that keep track of all retail purchases, and large panel data containing many parallel time series as the one used in Chapter 3.

The availability of powerful and cheap computers generated not only improvements in storage capacity, but also allowed researchers to move from estimation methods that could be evaluated analytically, to methods that could be numerically computed, and finally to estimation methods based on approximations computed via Monte Carlo simulations.

More specifically, Gourieroux and Monfort (2002) distinguish three stages in the evolution of econometrics. The first one was dominated by linear models and the least squares approach, which lead to estimators with analytical form.

The second period started with the introduction of numerical optimization algorithms in the 1970s, which allowed to obtain estimates and their estimated precision without knowing the analytical expression of the estimators. This is the period of the Maximum Likelihood (ML) approach, and of the Generalized Method of Moments (GMM), which are methods based on the optimization of a nonquadratic criterion function that has closed-form expression.

Nowadays, econometricians are able to deal with problems where the criterion function to be optimized does not have a tractable expression, and must be approximated. Typically the criterion functions have no closed-form solution because they contain high dimensional integrals that cannot be solved analytically.

A common example of this kind of problem is the estimation of dynamic latent variable (DLV) models, which uses variables that are not directly observable to model complex phenomena in a flexible way. Latent variables are recognized to be major components of the behavior of economic agents (McFadden, 1989), and are inherently dynamic for models including intertemporal optimization, rational expectations, error correction mechanisms, and state dependence. Some examples within this class are state-space models, unobserved components models, models with varying coefficients, limited dependent variables and truncated regression models, and switching regime models.

Inference in DLV models may be difficult because of multidimensional integrals over latent factors appearing in the probability density function of the observable variables. Without the density of the observables, analytic moments cannot be computed, complicating maximum likelihood, and moment-based estimation methods (Creel, 2008). In some particular cases such as linear state-space and Markov-switching models, these integrals can be solved analytically (see, e.g. Kalman, 1960 and Hamilton, 1989). In other cases, numerical methods can be used to evaluate the criterion function (see Kitagawa, 1987), or the objective function can be computed from approximated models as in Harvey et al. (1994).

However, in cases such as discrete choice models (see Liesenfeld and Richard, 2008b), some stochastic volatility models (see Liesenfeld and Richard, 2003), non-Markovian switching regime models (see Billio and Monfort, 1998), or nonlinear DSGE models (see Fernandez-Villaverde and Rubio-Ramirez, 2005), multidimensional integrals in the density of the observable cannot be computed analytically. Digital computers allow the estimation of these otherwise intractable models via the approximation of large dimensional integrals using MC simulations.

Metropolis and Ulam (1949) define the Monte Carlo method as a technique to approximate the solution to an analytical intractable problem by replacing it with a probabilistic one with the same solution, and solving the later problem using statistical sampling. According to Stern (1997), the generic MC integration problem is

$$I = \int_{\mathbb{R}^n} g(x) \cdot f(x) dx, \quad (1.1)$$

where x is a vector of random variable of dimension n with probability density function $f(x)$, and $g(x)$ is a general function. Note that equation (1.1) can be

1. INTRODUCTION

interpreted as the following expectation

$$I = E_f[g(x)], \quad (1.2)$$

where E_f denotes expectation with respect to the density function $f(x)$. If it is possible to draw random variables $\{x_i\}_{i=1}^N$ from $f(x)$, the probabilistic problem in (1.2) can be approximated by statistical sampling (1.2) as an average

$$I \approx \frac{1}{N} \sum_{i=1}^N g(x_i), \quad (1.3)$$

which, under certain conditions (see Geweke, 1989a), will converge to the true value of the integral in (1.1) with rapidity $N^{-\frac{1}{2}}$.

The possibility to approximate high dimensional integrals as in (1.1) is what gave rise to the “simulation revolution in Bayesian econometric inference” described by van Dijk (1999). The engineering breakthroughs in computer technology allowed the developments of algorithms that made possible to simulate draws from a posterior density without knowing its integrating constant, and let Bayesian statistics and econometrics flourish.

Simulation-based estimation allowed the statistical analysis of a much broader class of problems, and various simulation-based inference methods were developed and applied to DLV models. The simulated EM method was used e.g., by Shephard (1993) to estimate stochastic volatility models, simulated pseudo-maximum likelihood methods have been applied to disequilibrium models by Laroque and Salanie (1993), Albert and Chib (1993) developed Bayesian methods based on data augmentation principles to estimate DLV models. The indirect inference method proposed by Gouriéroux et al. (1993), and the method of simulated moments (see Duffie and Singleton, 1993) have also been used to estimate DLV models.

Billio and Monfort (1998) used simulated likelihood techniques based on Importance Sampling methods to estimate non-Markovian switching regimes models. Liesenfeld and Richard (2008b) and Liesenfeld and Richard (2003) used Efficient Importance Sampling to estimate nonlinear panel data models with latent factors, and stochastic volatility models.

Since Hammersley and Handscomb (1964), and Kloek and van Dijk (1978), *Importance Sampling* (IS) has become an important simulation tool. IS provides

a very elegant solution when it is hard to draw random variables from the density $f(x)$ in (1.1). The main idea is instead to draw samples from an auxiliary density $m(x)$, often called *importance sampler*, and to weight them to account for the differences between $m(x)$ and $f(x)$:

$$\begin{aligned}
 I &= \int_{\mathbb{R}^n} g(x) \cdot \frac{f(x)}{m(x)} \cdot m(x) dx, \\
 I &\approx \frac{1}{N} \sum_{i=1}^N g(x_i) \cdot w(x_i),
 \end{aligned} \tag{1.4}$$

where

$$w(x_i) = \frac{f(x)}{m(x)},$$

are the importance weights, and $\{x_i\}_{i=1}^N$ are now drawn from the importance sampler $m(x)$.

Besides providing a way to use MC simulation to solve problems where it is hard to sample from $f(x)$, IS can also be used to achieve variance reduction (Robert and Casella, 2005). Note that the MC method used in (1.1) samples from $f(x)$ and completely disregard its interaction with $g(x)$. This can be highly inefficient when large values of g are concentrated in a relative small part of the domain of f . However, the importance sampler $m(x)$ can be tailored in such a way that it concentrates the distribution of the sample points $\{x_i\}_{i=1}^N$ in parts of the region of integration where the integrand $g(x) \cdot f(x)$ in (1.1) has large values, instead of sampling x_i from f irrespectively of g . Sampling more often from the important parts of the target integrand can drastically reduce the MC sampling variance of the approximations.

van Dijk (1999) argues that because of the difficulty in finding mechanical multivariate importance samplers, importance sampling did not experience the same success of Markov Chain Monte Carlo (MCMC) methods. MCMC methods are algorithms for sampling from probability distributions based on a Markov chain that has the desired distribution as its stationary probability distribution, and most of these algorithms rely upon Gibbs Sampling and Metropolis-Hastings (MH) to construct the desired chain (see Chib, 2001). Gibbs procedures are based on the fact that *conditional* on some parameters one is often able to determine the functional form of the posterior in such a way that it is possible to draw from the

1. INTRODUCTION

conditional posterior distribution. The Metropolis-Hastings algorithm rests on the idea that although it is not possible to draw from the posterior, there is usually a dominating function of it from which it is easy to generate random drawings, and thus MH has some similarities with IS (see Liesenfeld and Richard, 2008a). However, van Dijk (1999) warns that naive use of MCMC methods often leads to misleading results. More specifically, mechanical construction of Markov chains using Gibbs and/or MH algorithms is possible even without knowing anything about the shape of the posterior distribution. Nevertheless, van Dijk (1999) shows that without a careful study of the shape of the posterior surface one cannot assure that the convergence conditions stated, e.g. in Geweke (1999), are satisfied.

Efficient Importance Sampling (EIS) is a method developed by Richard and Zhang (2007) to overcome the problem of finding mechanical multivariate importance samplers, and can be used to efficiently solve high dimensional integration problems appearing in the estimation of different dynamic latent variable models. The EIS method is a generic algorithm to construct efficient importance samplers searching within a class of parametric densities for the importance sampler delivering the best approximation to the target integrand $g(x) \cdot f(x)$. The procedure is based on a recursive sequence of auxiliary least squares regressions, which are designed to estimate the parameters of the importance sampling density inside the chosen parametric class that better approximate the integrand, minimizing the Monte Carlo variance of the integral approximation. As the EIS regressions cover the full support of the integrand, EIS samplers can be seen as global approximations to the target integrand, providing huge efficiency gains in terms of variance reduction.

Additionally, when the chosen class of parametric densities belong to the exponential family, the EIS regressions can be set in such a way that they are linear in the parameters of interest. Therefore, a sequence of standard Ordinary Least Squares (OLS) regressions can be used in the search for the parametrization of the importance sampler that better approximates the target integrand. This simplifies considerably the whole procedure, speeding up computations.

Liesenfeld and Richard (2008a) note that the sampling properties of both importance sampling and MCMC methods depend upon the adequacy of an auxiliary sampler as an approximation to a density kernel which needs to be

numerically integrated. Particularly, they argue that the problem of infinite variance often raised in importance sampling applications also applies to Metropolis-Hastings procedures. Given the similarities between the two approaches, Liesenfeld and Richard (2008a) show that EIS is able to facilitate the design, as well as to improve the sampling properties of proposal densities used in MCMC methods.

This work uses EIS to efficiently solve high dimensional integration problems appearing in the estimation of different dynamic latent variable models. Chapter 2 presents deterministic and Monte Carlo integration methods that are used throughout the thesis to approximate the solution of integrals like the one in (1.1). In Chapter 3, efficient high-dimensional integration problems arising in the likelihood evaluation of nonlinear panel data models with different kinds of unobserved components are solved in order to analyze the determinants and dynamics of current account reversals. Chapter 4 develops the EIS filter, a sequential Monte Carlo filter for applications involving nonlinear and non-Gaussian state space models. The filter is used to perform likelihood evaluation in nonlinear DSGE models. The Appendix to Chapter 4 presents comparisons of different nonlinear solution methods to DSGE models.

1.1 Outline

Deterministic methods of numerical integration use ideas from interpolation theory to approximate the target integrand using polynomials, and integrate them analytically. The exact solution to the polynomial integration is then used to approximate the unknown integral (see Davis and Rabinowitz, 1984).

Interpolation of univariate functions is usually easy and very accurate, and deterministic methods of integration inherit this characteristic. However, the extension to multivariate functions is problematic given the complexity of multidimensional spaces, implying that the number of nodes required to interpolate the integrand increases very fast with the dimensionality of the problem (see Stroud, 1971). This characterizes the *curse of dimensionality*, and is one of the main drawbacks of deterministic integration methods.

Monte Carlo integration methods have convergence rate of $N^{-\frac{1}{2}}$, which is not very fast when compared to deterministic univariate integration methods that can have rate of convergence of e^{-cN} , where c is a constant and N is the number of

1. INTRODUCTION

interpolation nodes. However, different from deterministic quadrature methods, the speed of convergence of MC methods does not depend on dimensionality of the integration problem at hand (Fishman, 1996). This is one of the main reasons why MC methods are so popular in solving multidimensional integrals.

In the present work, univariate integration problems will be typically approximated via deterministic quadrature rules, while multivariate problems will be approximated using Monte Carlo integration methods.

In Chapter 3, which was co-authored with Roman Liesenfeld and Jean-François Richard, the determinants and dynamics of current account reversals for a panel of developing and emerging countries will be analyzed controlling for alternative sources of persistence.

Likelihood evaluation of nonlinear panel data models with unobserved heterogeneity and dynamic error components is complicated by the fact that the computation of the event probabilities requires high-dimensional interdependent integration over these latent variables. The dimension of such integrals is typically given by the number of time periods (T), or if one allows for interaction between individual and time specific random effects by $T + N$, where N is the number of units in the panel. Thus likelihood estimation of such models typically relies upon Monte Carlo integration techniques.

Heckman (1981) argues that there are two distinct possible sources of serial dependence which ought to be taken into account when analyzing discrete variable panel data model: state dependence and unobserved heterogeneity across units. In the context of current account reversals, state dependence would reflect the possibility that past reversals could affect the probability of another reversal, and unobserved heterogeneity would reflect differences in institutional, political or relevant economic factors across countries which cannot be controlled for. Therefore, the starting point is a panel probit model with state dependence and random heterogeneity that requires only univariate integration, and the approach of Butler and Moffitt (1982) is used to estimate the model using deterministic integration methods.

However, as argued, e.g., by Hyslop (1999), serial dependence could also be transitory resulting from autocorrelated country specific errors reflecting the effects of omitted serially correlated macroeconomic factors or serially correlated

country-specific shocks like regional conflicts, uncertainty about government transition, and political changes. To analyze the robustness of the initial panel probit model with state dependence and random heterogeneity, a model allowing for correlated idiosyncratic error components is estimated.

Likelihood evaluation of panel probit models with unobserved heterogeneity and autocorrelated error components requires approximating integrals with dimension given by the number of time periods (T). As quadrature methods may become infeasible for realistic values of T , various MC procedures have been proposed for the evaluation of such high-dimensional integrals. The most popular among those is the importance sampling procedure known as GHK (see, eg., Geweke and Keane, 2001) which has been applied to the estimation of dynamic panel probit models, e.g., by Falcetti and Tudela (2006). In Chapter 3, EIS is going to be used to improve upon the GHK importance sampler with remarkable success in reducing numerical errors associated with likelihood approximations.

Serially correlated time random effects is an interesting extension in trying to account for possible dynamic spillover effects of current account crises, or global shocks such as oil or commodity price shocks. In particular, following the financial turbulences of the 1990s, which rapidly spread across emerging countries, it is recognized that spillover effects are important. A crisis in one economy can also affect the fundamentals of other countries through trade links and currency devaluations. Trading partners of a country in which a financial crisis has induced a sharp currency depreciation could experience a deterioration of their trade balance and current account resulting from a decline in exports and an increase in imports. In the words of the former Managing Director of the IMF: “from the viewpoint of the international system, the devaluations in Asia will lead to large current account surpluses in those countries, damaging the competitive position of other countries and requiring them to run current account deficit” Fisher (1998). When country specific and time random effects are allowed to interact, the likelihood function requires the approximation of $T + N$ dimensional integrals. EIS is used to efficiently solve these high-dimensional integrals taking into account the *a posteriori* dependence between the two random effects.

Chapter 4, which was co-authored with David DeJong, Dharmarajan Hariharan, Roman Liesenfeld, and Jean-François Richard, is concerned with likelihood evaluation and filtering in nonlinear state-space models. Filtering is important

1. INTRODUCTION

in the time series analysis of state-space models not only as a means to infer the state of the system, but also as a means to compute the likelihood function and to estimate the parameters of these models. Likelihood evaluation and filtering in applications involving state-space models require the calculation of integrals over unobservable state variables in order to compute unknown densities in the filtering process (see Kitagawa, 1987). When models are linear and stochastic processes are Gaussian, required integrals can be calculated analytically via the Kalman filter, and all unknown densities have close-form expression. Departures entail integrals that must be approximated numerically, either via deterministic quadrature methods or via MC integration. Since deterministic methods become unfeasible when the number of states is large, sequential Monte Carlo methods like the standard particle filter have become the standard tool in nonlinear filtering (see Doucet et al., 2001).

Sargent (1989) demonstrated the mapping of DSGE models into a state-space representations, and showed how to perform likelihood-based analysis using a linear approximation to the original model using the Kalman filter. Likelihood-based inference in nonlinear DSGE models has been implemented in the literature using the particle filter, eg., by Fernandez-Villaverde and Rubio-Ramirez (2005, 2007), An and Schorfheide (2007), and Amisano and Tristani (2007). Although conceptually simple, the particle filter uses discrete approximation to unknown densities in the filtering process, which translates into spurious likelihood discontinuities. Moreover, the support upon which particle filter's density approximations are based is not adapted, generating sample impoverishment when the observable variables at t are very informative with respect to the states.

In order to overcome the problems of the particle filter, an efficient filtering procedure for nonlinear state space models based on continuous and fully adapted importance samplers constructed via EIS is presented in Chapter 4. The EIS filter is able to precisely determine the importance region of the target densities, implying huge time and efficiency gains, and allowing reliable likelihood evaluation of nonlinear state space models.

The shortcomings of the particle filter and their effects on the numerical accuracy of likelihood evaluations, as well as the ability of the EIS filter to overcome them are illustrated in Chapter 4 with Monte Carlo experiments based on a two-state Real Business Cycle (RBC) model, and on a six-state Small Open Economy

(SOE) model. Two data sets for each models are analyzed: an artificial data set generated from a known model parameterization; and a corresponding real data set.

The RBC model used is the one constructed and estimated using the particle filter by Fernandez-Villaverde and Rubio-Ramirez (2005). Both data sets used were the same as in Fernandez-Villaverde and Rubio-Ramirez (2005), and each of them poses a distinct challenge to efficient filtering. In the artificial data set, the standard deviations of the measurement errors are small relative to shocks to the unobserved states, which can lead to problems associated with sample impoverishment. In the real data set, the investment series contains two outliers. Outliers can induce bias in likelihood estimates associated with the particle filter.

The small-open-economy (SOE) model was patterned after those considered, e.g., by Mendoza (1991) and Schmitt-Grohe and Uribe (2003), but extended to include six state variables to show that the EIS filter does not suffer badly from the curse of dimensionality. In this application neither data set contains an outlier observation, nor standard deviations of measurement errors are small. As opposed to the applications involving the RBC model, variances of measurement errors are closely comparable across data sets. Instead, differences stem primarily from differences in the volatility and persistence of the model's structural shocks. In particular, with the model parameterization associated with the artificial data set calibrated to annual data, structural shocks are far less persistent, and generally more volatile than in the parameterization associated with the real data set estimated using quarterly observations. The upshot is that in working with the actual data, the state variables are relatively easy to track, and in general the construction of likelihood approximations is less problematic.

Before one turns to the likelihood evaluation of nonlinear DSGE models, it is necessary to decide upon a solution method that will be used to obtain the nonlinear approximations to the policy functions used in the computation of the likelihood. Aruoba et al. (2006) provide some guidance on this choice in the context of an RBC model, but because relative performance of different solution methods are model dependent, the Appendix to Chapter 4 compares different nonlinear solution methods using the stochastic growth model from Taylor and Uhlig (1990).

1. INTRODUCTION

If the solution method chosen is based on local approximations around the non-stochastic steady state, the computation of agent's expectations about the future values of the model variables is avoided. However, expectations must be computed numerically if the interest is on global approximations to the policy functions. Furthermore, some nonlinear methods construct approximations to the policy functions requiring that the approximation error is zero on a weighted integral sense, requiring again numerical solution of integrals. Since these integrals are of low dimensions in the case of the stochastic growth model, different deterministic quadrature rules are used to approximate them.

Chapter 2

Numerical Integration

In economic and econometric applications, one is often interested in the solution to an integral of a real-valued function g with respect to a positive weight function f over $D \subseteq \mathbb{R}^n$

$$I = \int_D g(x) \cdot f(x) dx, \quad (2.1)$$

where $\int_D f(x) dx = 1$. In some cases, the weight function f is the identity $f(x) \equiv 1, \forall x \in D$, in which case the integral in (2.1) would represent the area under the function g . When $f(x)$ is the probability density function (pdf) of a continuous random variable x with support D , (2.1) represents the expectation of $g(x)$.

Numerical integration methods may be used to approximate the solution to (2.1), because either the antiderivative of $g(x)$ might not possess an analytical solution, or because $g(x)$ is only known at certain points, or even because it may be easier to approximate (2.1) numerically than to solve it analytically. Integrals like the one in (2.1) are often approximated as a weighted sum of function values:

$$\int_D g(x) \cdot f(x) dx \approx \sum_{i=0}^n w_i \cdot g(x_i), \quad (2.2)$$

where w_i are the weights, and x_i are the abscissa points.

This chapter presents deterministic and Monte Carlo (stochastic) methods to approximate integrals as in (2.2). The main goals are to review some concepts and methods, define the notation and the terminology that will be used in the coming chapters, and to describe the advantages and disadvantages of the different approaches. Deterministic methods of integration yield precise results with

2. NUMERICAL INTEGRATION

relative few integrand evaluations in one dimension, but their computational burden grow exponentially with the dimension of integration, what is often referred to as the *curse of dimensionality* after Bellman (1961). Monte Carlo methods on the other hand have slower convergence rates in one dimension, but these rates are independent of the dimension of the problem at hand, which make them more suitable for multidimensional integration.

A more complete discussion of deterministic methods of integration can be found in Davis and Rabinowitz (1984), whereas Stroud and Secrest (1966) focus on Gaussian quadrature methods, and Stroud (1971) on deterministic methods for the approximate calculation of multiple integrals. A classical presentation of Monte Carlo methods is Hammersley and Handscomb (1964), while an up to date throughout presentation can be found on Fishman (1996). Robert and Casella (2005) present many Monte Carlo based statistical methods that are important in statistics and econometrics.

2.1 Deterministic Methods of Integration

The study of deterministic methods of numerical integration dates from antiquity, when Archimedes used regular polygons to find an upper and a lower bound for π , and continued to be an active area of research over the centuries, with names like Kepler, Newton, Euler, Gauss and other contributing to the field (see Davis and Rabinowitz, 1984).

Deterministic quadrature methods use ideas from mathematical interpolation theory to evaluate the integrand at a finite number of points, use this information to construct a polynomial approximation to it, integrate this approximation exactly, and uses it to approximate (2.1). Therefore, it may be interesting to review some concepts of polynomial interpolation theory that are important to understand these methods.

Davis (1963) regards polynomial interpolation as a variation on two theorems: the interpolation theorem, and Weierstrass' approximation theorem. Let \mathcal{P}_n be the class of polynomials of degree $\leq n$, $\mathcal{C}^l[a, b]$ be the class of all functions with continuous l^{th} derivative on $[a, b]$, and $\mathcal{C}^\infty[a, b]$ the class of all continuous functions on $[a, b]$,

2.1 Deterministic Methods of Integration

Theorem 2.1.1 *Given $n + 1$ distinct points x_0, x_1, \dots, x_n and $n + 1$ values g_0, g_1, \dots, g_n . There exists a unique polynomial $P_n(x) \in \mathcal{P}_n$ for which*

$$P_n(x_i) = g_i, \quad i = 0, 1, \dots, n. \quad (2.3)$$

and

$$P_n(x) = g_0 \cdot L_{n,0}(x) + g_1 \cdot L_{n,1}(x) + \dots + g_n \cdot L_{n,n}(x) \quad (2.4)$$

where

$$L_{n,k}(x) = \prod_{\substack{i=0 \\ i \neq k}}^n \frac{(x - x_i)}{(x_k - x_i)}. \quad (2.5)$$

P_n is called the n^{th} Lagrange interpolating polynomial.

The Lagrange interpolating polynomial is not the only representation possible of a n^{th} polynomial, but as there exists only a unique polynomial satisfying (2.3), all other representations of P_n can always be written in the form of (2.5) ¹.

Theorem 2.1.1 states that it is possible to fit a n degree polynomial to $n + 1$ points, which means that a straight line is determined by two points, a parabola by three, a cubic by four, and so on. Although simple, much of numerical analysis rests upon this theorem. It suggests that polynomials might be a good choice to approximate unknown functions, whose values at some arbitrary points are known. The other important theorem of mathematical interpolation theory shows that this conjecture is right at least for continuous functions.

Theorem 2.1.2 (*Weierstrass*) *Let $g(x) \in \mathcal{C}^\infty[a, b]$. Given an $\epsilon > 0$ it is possible to find a polynomial $P_n(x)$ of sufficiently high degree for which*

$$\|g - P_n\|_\infty \equiv \sup_{x \in [a,b]} |g(x) - P_n(x)| \leq \epsilon, \quad a \leq x \leq b. \quad (2.6)$$

Weierstrass' theorem asserts the possibility of approximating any continuous real-valued function g defined on a bounded interval of the real line to any degree of accuracy using a polynomial, which provides a strong motivation for using

¹According to Acton (1997) "Lagrange polynomials are praised for its analytic utility and beauty, but deplored for its numerical properties".

2. NUMERICAL INTEGRATION

polynomials to approximate this kind of functions, but does not say anything about how to construct a good polynomial approximant ².

Different polynomial approximant is exactly what differentiates the many deterministic quadrature rules, and determines their degree of precision. It is usually said that a formula has degree m if it is exact whenever $g(x)$ is a polynomial of degree $\leq m$, and is not exact for at least one polynomial of degree $> m$.

It is important to note that deterministic quadrature methods approximate integrands using polynomials, which are intimately related to continuous functions through theorem 2.1.2. Therefore, the quality of the integral approximation resulting from these methods depends on the degree of smoothness of the target integrand, where smoothness is related to continuity, number of continuous derivatives and their magnitude, so that a function belonging to $\mathcal{C}^l[a, b]$ is said to have l^{th} order smoothness.

The smoother the function, the closer the approximation, and the faster the convergence of a sequence of approximations. More specifically, a sequence of deterministic quadrature rules of degree m will converge to the true value of the integrand with rapidity at least N^{-k} , where $k = \min(l, m + 1)$, and N is the number of function evaluations, which implies that specifying $m \geq l - 1$ allows the error to go to zero about as fast as N^{-l} .

2.1.1 Univariate Quadrature Methods

This subsection presents univariate deterministic methods of integration. The theory of numerical integration for functions of one variable is much better developed than the one for multivariate functions, the reason being that one-dimensional spaces are much simpler geometrically. Furthermore, multivariate rules are often based on tensor product extensions of univariate rules, so it might be interesting to focus on univariate rules. In section 2.1.2 generalizations to multivariate case will be discussed.

²See Davis (1963) for proofs of theorems 2.1.1, and 2.1.2.

2.1 Deterministic Methods of Integration

Using theorems 2.1.1 and 2.1.2 we can rewrite (2.2) for the univariate case as

$$\begin{aligned}
 \int_a^b g(x)f(x) dx &\approx \int_a^b P_n(x)f(x) dx, \\
 &\approx \int_a^b \sum_{i=0}^n g(x_i)f(x)L_{n,i}(x) dx, \\
 &\approx \sum_{i=0}^n \left[\int_a^b L_{n,i}(x)f(x) dx \right] g(x_i), \\
 &\approx \sum_{i=0}^n w_i g(x_i),
 \end{aligned} \tag{2.7}$$

where

$$w_i = \int_a^b L_{n,i}(x)f(x) dx, \tag{2.8}$$

which shows how the quadrature weights are computed, and how they depend on the weight function $f(x)$.

2.1.1.1 Newton-Cotes quadratures

The classical and still useful Newton-Cotes method of numerical integration was developed in the beginning of the eighteenth century independently by Isaac Newton and Roger Cotes (see Chabert, 1999). They are designed based on piecewise-polynomial approximations to the integrand of (2.1), and the weight used is $f(x) \equiv 1$. Examples of Newton-Cotes quadratures are the midpoint rule, trapezoid rule, Simpson's rule, and Bode's rule. Here only Simpson's rule will be presented because of its wide spread use in obtaining approximate integrals. For other types of Newton-Cotes rules see Press et al. (2007). Simpson's rule uses a piecewise-quadratic approximation to $g(x)$ with three abscissa points given by: $x_0 = a$, $x_1 = \frac{1}{2}(a + b)$, and $x_2 = b$. Therefore, its weights are:

$$\begin{aligned}
 w_0 &= \int_a^b L_{2,0}(x) dx = \int_{x_0}^{x_2} \frac{(x - x_1) \cdot (x - x_2)}{(x_0 - x_1) \cdot (x_0 - x_2)} dx = \frac{1}{6}(b - a), \\
 w_1 &= \int_a^b L_{2,1}(x) dx = \int_{x_0}^{x_2} \frac{(x - x_0) \cdot (x - x_2)}{(x_1 - x_0) \cdot (x_1 - x_2)} dx = \frac{2}{3}(b - a), \\
 w_2 &= \int_a^b L_{2,2}(x) dx = \int_{x_0}^{x_2} \frac{(x - x_0) \cdot (x - x_1)}{(x_2 - x_0) \cdot (x_2 - x_1)} dx = \frac{1}{6}(b - a),
 \end{aligned}$$

2. NUMERICAL INTEGRATION

and the approximate integral is

Simpson's Rule

$$\int_a^b g(x) dx = \frac{b-a}{6} [g(x_0) + 4g(x_1) + g(x_2)] - \frac{(b-a)^5}{2880} g^{(4)}(\xi), \quad (2.9)$$

where $g^{(4)}(\xi)$ denotes the fourth derivative of g evaluated at ξ , and $\xi \in [a, b]$. From theorem 2.1.1 one would expect that Simpson's rule would be exact for polynomials of degree 2 or less as it uses 3 abscissas, but because of a cancelation of coefficients due to the left-right symmetry of the grid points, this three-point formula is of degree 3.

Simpson's rule is most frequently applied as a sequence. This is the compound, or composite form, in which the interval $[a, b]$ is divided into equal subintervals, and Simpson's rule is applied to each one of them. Let $a = x_0 < x_1 < \dots < x_{2n-1} < x_{2n} = b$ be a sequence of equally spaced points in $[a, b]$, and $h = x_{i+1} - x_i$ for $i = 0, \dots, 2n - 1$. Then the compound Simpson's rule is:

$$\int_a^b g(x) dx = \frac{h}{3} \{g(x_0) + 4[g(x_1) + g(x_3) + \dots + g(x_{2n-1})] + 2[g(x_2) + g(x_4) + \dots + g(x_{2n-2})] + g(x_{2n})\} + E_n, \quad (2.10)$$

where $E_n = -\frac{nh^5}{90} g^{(4)}(\xi)$, and $\xi \in [a, b]$. The compound Simpson's rule can be rewritten as in (2.2), where $w_1 = w_{2n} = \frac{h}{3}$, and $w_i = \frac{4h}{3}$ if i is even or $w_i = \frac{2h}{3}$ if i is odd. For functions that have four continuous derivatives, this rule converges to the true value of the integral with a rate at worst equal to N^{-4} , where $N = 2n$ is the number of subdivisions of $[a, b]$. This implies that ten subintervals should be enough to secure four decimals of precision.

2.1.1.2 Gaussian quadratures

Theorem 2.1.1 asserts that it is possible to fit a n degree polynomial to $n + 1$ points, but because of the position of the abscissa points, Simpson's rule was able to do better than that, and to use 3 points to perfectly fit a polynomial of degree 3. Gaussian quadrature methods look for the highest degree formula that can be obtained with n points. They use (2.2) to approximate integrals, but unlike Newton-Cotes rules, not only the weights w_i are carefully chosen, but also the abscissa points x_i , which provides twice the number of degrees of freedom. It

2.1 Deterministic Methods of Integration

turns out that it is possible to construct Gaussian quadrature formulas whose order is twice that of Newton-Cotes rules with the same number of function evaluations. All Gaussian quadratures are based on the following theorem

Theorem 2.1.3 *Let $P_n(x)$ be an orthogonal polynomial of degree n , such that:*

$$\int_a^b P_n(x) \cdot f(x) \cdot x^k dx = 0, \quad (2.11)$$

where k is an integer on $[0, n - 1]$, and $f(x)$ is a weight function. Let $\{x_i\}_{i=1}^n$ be the n roots of $P_n(x)$, then it is possible to find $\{w_i\}_{i=1}^n$ such that

$$\int_a^b g(x) \cdot f(x) dx \approx \sum_{i=1}^n w_i \cdot g(x_i), \quad (2.12)$$

and (2.12) holds with equality if $g(x) \in \mathcal{P}_{2n-1}$.

The basic interpolation theorem 2.1.1 says that it is always possible to fit a n degree polynomial to a set of $n + 1$ points, but theorem 2.1.3 treats $\{x_i\}_{i=1}^n$ and $\{w_i\}_{i=1}^n$ as $2n$ unknowns, and craftily chooses them so that it is possible to use a n -degree polynomial to perfectly fit polynomials of the class \mathcal{P}_{2n-1} . This result is intimately related to the orthogonal polynomials generated by the weight function $f(x)$, and Gaussian rules will differ from one another in the way $f(x)$ and the applicable orthogonal polynomial are defined³.

The weights of Gaussian quadratures are chosen so as to satisfy the following $2n$ moment-matching conditions:

$$\int_a^b x^k \cdot f(x) dx = \sum_{i=1}^n w_i \cdot x_i^k, \quad k = 0, \dots, 2n - 1. \quad (2.13)$$

Gaussian formulas are very accurate in practice if the integrand is analytic in the interval of integration, and the error can be shown to go to zero with rapidity e^{-cn} for some constant c (see Haber, 1970). Here only two elements of the class of Gaussian quadratures that are going to be used in later chapters will be presented, for a more complete discussion of different Gaussian rules see Stroud and Secrest (1966), Davis and Rabinowitz (1984), and Press et al. (2007).

Gauss-Legendre quadrature is a special case of Gaussian quadratures where the weight function is the identity $f(x) \equiv 1$, and thus it is suitable for computing the area under a curve.

³See Stroud and Secrest (1966) for a proof of Theorem 2.1.3

2. NUMERICAL INTEGRATION

Gauss-Legendre Quadrature

$$\int_{-1}^1 g(x) dx = \sum_{i=1}^n w_i \cdot g(x_i) + \frac{2^{2n+1}(n!)^4}{(2n+1)!(2n)!} \cdot \frac{f^{(2n)}(\xi)}{(2n)!}, \quad (2.14)$$

where $-1 \leq \xi \leq 1$, the abscissas $\{x_i\}_{i=1}^n$ are the zeros of the n^{th} order Legendre polynomial, and the weights are given by

$$w_i = \int_{-1}^1 \prod_{\substack{i=0 \\ i \neq j}}^n \frac{(x - x_i)}{(x_j - x_i)}. \quad (2.15)$$

Equation (2.14) is defined over the interval $[-1, 1]$, but a simple linear change of variables enable its use for general integrals

$$\int_a^b g(x) dx \approx \frac{b-a}{2} \sum_{i=1}^n w_i \cdot g\left(\frac{(b-a)z_i + b + a}{2}\right). \quad (2.16)$$

Press et al. (2007) and Mirana and Fackler (2002) provide routines to compute the weights and abscissas for Gauss-Legendre quadrature, while Stroud and Secrest (1966) and Judd (1998) present tables with weights and abscissas for different values of n .

Gaussian quadratures are normally of rapid convergence, and the Gauss-Legendre rule inherits this property with the error being bounded above by $\pi 4^{-n} M$, where

$$M = \sup_m \left[\max_{-1 \leq x \leq 1} \frac{g^{(m)}(x)}{m!} \right],$$

which means that when M is finite, as for analytic functions, Gauss-Legendre quadrature converges exponentially as n increases (see Judd, 1998). This implies that Gauss-Legendre quadrature converges much faster, and thus are much better than Newton-Cotes formulas when $g \in \mathcal{C}^\infty$.

Another very important quadrature rule of the Gauss-type is the Gauss-Hermite quadrature. It is based on a weight function given by e^{-x^2} , and is a natural choice to be used in connection with normal random variables.

Gauss-Hermite Quadrature

$$\int_{\mathbb{R}} g(x) \cdot e^{-x^2} dx = \sum_{i=1}^n w_i \cdot g(x_i) + \frac{n! \sqrt{\pi}}{2^n} \cdot \frac{f^{(2n)}(\xi)}{(2n)!}, \quad (2.17)$$

2.1 Deterministic Methods of Integration

where $\xi \in \mathbb{R}$.

The following identity is useful in applying the Gauss-Hermite formula to compute e.g., expectations of functions of a normal random variable $x \sim N(\mu, \sigma^2)$

$$\int_{\mathbb{R}} g(x) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{\mathbb{R}} g(\sqrt{2}\sigma x + \mu) e^{-x^2} \sqrt{2}\sigma dx. \quad (2.18)$$

Using (2.18) and a linear change of variables $z = \frac{(x-\mu)}{\sqrt{2}\sigma}$, yields

$$\begin{aligned} E[g(x)] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} g(x) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &\approx \frac{1}{\sqrt{\pi}} \sum_{i=1}^n w_i g(\sqrt{2}\sigma z_i + \mu), \end{aligned} \quad (2.19)$$

where the Gauss-Hermite weights $\{w_i\}_{i=1}^n$ are obtained from (2.13), and the abscissas $\{z_i\}_{i=1}^n$ are the roots of the n^{th} order Hermite orthogonal polynomial. Press et al. (2007) and Mirana and Fackler (2002) also provide routines to compute the weights and abscissas of Gauss-Hermite quadrature, while Stroud and Secrest (1966) and Judd (1998) present tables with weights and abscissas for different values of n .

The question of whether to use a Newton-Cotes type rule or a Gaussian rule depends very much on the integrand being evaluated. If the integrand is not very smooth, belonging to \mathcal{C}^l but not to \mathcal{C}^{l+1} , for a fairly low value of l , the use of a composite Newton-Cotes formula of degree l will take advantage of all the smoothness the integrand has to offer. But if the integrand is analytic or belongs to \mathcal{C}^k , for a very high value of k , the quadrature rule should make use of all this smoothness, and then the Gaussian formulas often give the best results.

2.1.2 Multivariate Quadrature Methods

Univariate quadrature methods are well established in the literature, but their extension to multivariate cases are not straight forward due largely to the greater complexity of multidimensional spaces. All finite line segments in the one-dimensional Euclidean space E^1 are equivalent under an affine transformation, implying that formulas for a given univariate interval can be applied to any other interval in E^1 . For multidimensional spaces on the other hand, there are infinitely

2. NUMERICAL INTEGRATION

many distinct regions in E^n which are not equivalent under affine transformations. This means that integration rules for squares, circles, and triangles are all different from one another.

Multivariate integrals can sometimes be written as univariate integrals when some dimensions have analytical solutions, or after a suitable change of variables, but that is normally not the case in economics or econometrics (see Geweke, 1996).

A common approach to extend univariate rules to the multivariate case is to use the tensor product using the method of separation of variables. Consider the integral

$$\int_{\mathbb{R}^n} \cdots \int w(x_1, \dots, x_n) \cdot x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \dots \cdot x_n^{\alpha_n} dx_1 \dots dx_n. \quad (2.20)$$

If it is possible to find a possibly nonlinear transformation

$$\begin{aligned} x_1 &= x_1(z_1, \dots, z_n) \\ &\vdots \\ x_n &= x_n(z_1, \dots, z_n) \end{aligned}$$

which transforms (2.20) into a product of n univariate integrals, and if suitable formulas to approximate those single integrals are known, then these formulas can be combined to give a formula for \mathbb{R}^n . For n -cubes, n -spheres, and n -simplexes it is possible to construct multivariate formulas based on the product of n univariate integration rules (see Stroud, 1971). If the one-dimension formulas have n abscissa and degree of precision m , the resulting product rule will usually contain n^m points and degree of precision m .⁴

The fact that the number of points increase very rapidly as the dimensionality of the problem increases is maybe the most undesired property of deterministic multidimensional integration methods. While in one dimension the error is of the order N^{-l} , in s dimensions this error is only of the order $N^{-\frac{l}{s}}$, where again N denotes the total number of function evaluations, and l the degree of smoothness of the integrand. Although this gives only an upper bound on the error, Bahvalov's theorem states that there will be always a function g of degree of smoothness l for which any N -point deterministic formula will have error greater than $K(l)N^{-\frac{l}{s}}$

⁴Bungartz and Dirnstorfer (2003) use sparse grids to alleviate the curse of dimensionality in multivariate integration problems, generating multidimensional formulas with less points.

(see Haber, 1970). Which shows that deterministic numerical integration of functions of several variables requires a greater calculation effort than is required for univariate functions.

In Chapter 3, one-dimensional integrals will be approximated using deterministic quadratures with 20 abscissas, and similar multidimensional integrals in 72 variables using Monte Carlo methods. To integrate this kind of high-dimensional integrals using quadrature methods based on the product rule, one would need to evaluate the target integrand at $20^{72} \approx 4.7e + 93$ abscissa points, making them unfeasible for this kind of problems.

2.2 Monte Carlo Integration

The earliest use of statistical sampling to approximate integrals is believed to be that of Comte de Buffon. In 1777 he performed an experiment in which a needle was dropped many times onto a ruled board with equidistant parallel lines, and showed that the probability of the needle intersecting one of the lines is $\frac{2L}{\pi d}$, where L is the length of the needle and $d > L$ is the distance between the two lines. Later on, Laplace argued that this method could be used as a means of estimating π (see Katos and Whitlock, 2008). The main drawback of the method of Comte de Buffon is that it is rather slow, and it was not before the advent of electronic computers that its use became widespread.

During the 1940s, Stanislaw Ulam, at that time a scientist at Los Alamos laboratories, thought about an experiment similar to the one from Comte de Buffon. His question was about the chances that a Canfield solitaire card game laid out with 52 cards would come out successfully. He spent a lot of time trying to solve this problem using combinatorial calculus, just to realize that a more practical method would be to lay out the cards hundred times and observe the number of successful plays. At that time, the first electronic computer - the ENIAC - was being constructed at the University of Pennsylvania, and such an experiment was already possible to envisage (see Eckhardt, 1987; Metropolis, 1987).

Ulam realized that this methodology could also be used to solve many problems of mathematical physics that the scientists at Los Alamos were trying to solve. In 1946 he described these ideas to John von Neumann, and they began to

2. NUMERICAL INTEGRATION

plan calculations to solve problems of neutron diffusion using this method. The term Monte Carlo methods is often used to describe this rebirth of statistical sampling after the development of electronic computers ⁵. The term first appeared in Metropolis and Ulam (1949), who described it as a method to approximate solution of analytical intractable problems by replacing the original problem by a probabilistic one with the same solution, and then investigating the latter using statistical experiments ⁶.

When the dimensionality of the problem at hand is high, deterministic integration methods based on tensor product rules are likely to become impractical. Monte Carlo methods on the other hand, can be used for computing integrals accurately using a moderate number of abscissas. These methods are motivated by the law of large numbers, and by the central limit theorem. Let $\{x_i\}_{i=1}^N$ be an independent and identically distributed (i.i.d.) sample from a population, and let

$$\bar{x}_N = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{and} \quad s_N^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x}_N)^2.$$

Given that the population has finite first moment, then $E[\bar{x}_N] = E[x]$, the strong law of large numbers states that

$$\bar{x}_N \xrightarrow{a.s.} E[x]. \tag{2.21}$$

If the population also has finite variance σ^2 , then the central limit theorem establishes that

$$\sqrt{N} (\bar{x}_N - E[x]) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \tag{2.22}$$

and from the strong law of large numbers

$$s_N^2 \xrightarrow{a.s.} \sigma^2. \tag{2.23}$$

⁵Nicholas Metropolis suggested this name in reference to an uncle of Ulam, that would borrow money from his relatives to go to Monte Carlo (see Metropolis, 1987)

⁶An interesting reference on the origins of Monte Carlo methods is the special issue of *Los Alamos Science* published in memory of Stanislaw Ulam.

2.2.1 Classical Monte Carlo Integration

The idea of Monte Carlo integration is to evaluate the integral

$$I = E_f[g(x)] = \int_{\mathbb{R}^n} g(x) \cdot f(x) dx, \quad (2.24)$$

using a random sample (x_1, \dots, x_N) generated from the density function f to approximate (2.24) by empirical average

$$\bar{I}_N = \sum_{i=1}^N g(x_i), \quad (2.25)$$

where, differently from the earlier methods, the approximation \bar{I}_N itself is a random variable, and its primary source of error is not anymore due to numerical round-off, but due to the fact that only a finite sample can be taken. Thus the only assertion that can be made with certainty is that \bar{I}_N will lie somewhere between the maximum and the minimum value of g , as nothing prevents all x_i to fall into a particular small subregion of \mathbb{R}^n in a single sampling. However, because \bar{I}_N is a random variable, there is the possibility to make probabilistic statements about the accuracy of the Monte Carlo approximation \bar{I}_N . In fact, if $\int_{\mathbb{R}^n} g^2(x) f(x) dx$ is absolutely convergent, the variance of \bar{I}_N is given by

$$\text{var} [\bar{I}_N] = \frac{1}{N} \sigma^2, \quad (2.26)$$

where

$$\sigma^2 = \text{var}[g(x)] = \int_{\mathbb{R}^n} [g(x) - I]^2 \cdot f(x) dx, \quad (2.27)$$

and (2.26) can be estimated from the random sample (x_1, \dots, x_N) through

$$s_N^2 = \frac{1}{N(N-1)} \sum_{i=1}^N [g(x_i) - \bar{I}_N]^2. \quad (2.28)$$

From (2.21) it is clear that the Monte Carlo estimate in (2.25) converges almost surely to $E_f[g(x)]$. Furthermore, from (2.22)

$$\sqrt{N}(\bar{I}_N - I) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad (2.29)$$

2. NUMERICAL INTEGRATION

which shows that $\bar{I}_N - I$ is approximately normally distributed when N is large. Thus the probability that $|\bar{I}_N - I| > \epsilon$, where ϵ is any fixed number, can be assessed in terms of the ratio of ϵ to σ , for example

$$\text{Prob} \left(|\bar{I}_N - I| > 2\sigma N^{-\frac{1}{2}} \right) = 0.0455. \quad (2.30)$$

Note from (2.29) that approximations based on Monte Carlo integration converge to the true value of the integral with rapidity $N^{-\frac{1}{2}}$. In comparison with the convergence rates of univariate Gaussian quadratures this is not very rapid, but differently from the convergence of deterministic integration methods, the convergence of Monte Carlo methods does not depend neither on the dimensionality of the problem, nor on the degree of smoothness of the integrand. Moreover, the Monte Carlo method apply more broadly to bounded and unbounded functions, continuous and discontinuous functions, requiring only that $\int_{\mathbb{R}^n} g^2(x) f(x) dx < \infty$.

As noted in Section 2.1.2, the existence of infinitely affine inequivalent regions in all spaces with dimension greater than one implies the need of a different multivariate deterministic integration formula for each of those regions, while the formula in (2.25) may be used for any integration region with known volume.

Equation 2.28 shows another advantage of the Monte Carlo method, namely that it is possible to estimate the error of an integral approximation via the construction of confidence intervals around the approximation using generated data, whereas to determine the error from deterministic methods of integration is considerably more involved, and one has to know properties of the integrand such as its norm, or the norm of its derivatives.

The implementation of Monte Carlo methods depends on the possibility to generate sequences of random variables. In practice, a deterministic sequence of pseudorandom numbers is used in place of the random one, because methods to generate random numbers from physical processes are either too slow, or inconvenient.

Since pseudorandom numbers are generated by deterministic formulas, there are some philosophical objections to the use of the theorems of probability theory to these sequences of numbers, no matter how random they may seem to be (see Judd, 1998, chap. 8). When dealing with Monte Carlo approximations obtained using pseudorandom numbers, it is clearly not valid to use error estimates and

2.2 Monte Carlo Integration

confidence intervals based on the probability that a given sequence (x_1, \dots, x_N) lies in a certain set. This notion is based on the assumption that the sequence was *randomly* sampled from a probability space, but in the case of pseudorandom numbers it is known that the sequence is often taken from *a single point* of the sample space, which determines the seed of the deterministic generator. Nevertheless, this is what is customarily done and with remarkable success.

However, it should be realized that it is meaningless to ask whether a sequence of numbers is “truly random”, since it is impossible to prove if any particular process or device is random, or if it just seems random in light of the investigator’s ignorance. Furthermore, Whittle (1983) argues that probabilistic effects enter an econometric model typically to represent a host of minor deterministic effects which cannot be treated. Geweke (1996) makes a statement in favor of pseudorandom numbers using a similar idea. He argues that the assumption about the possibility to generate random numbers constitutes a model or idealization of what actually occurs, and compares this with the role played by the assumption of randomness in economic theory, or in the development of methods of statistical inference in econometrics. Pseudorandom number generators produce sequences of numbers that are in fact deterministic, but for which the assumption of i.i.d. is a model. The fact that the algorithm that produce a given observed sequence of pseudorandom numbers is known, implies that the data generating process is also known, which allows the analytical study of the properties of that sequence. On the other hand, real world events corresponding to the assumption of randomness in economic models have unknown data generating processes (that might be deterministic), precluding such a study of their analytical properties. In this sense, Geweke (1996) understands that the adequacy or inadequacy of the stochastic independence as a model for the pseudorandom number sequences is at least in a surer footing than is this assumption as a model in economic or econometric theory.

Algorithms to generate uniform pseudorandom numbers can be found in Fishman (1996), Press et al. (2007), or Geweke (1989a). Most other generating methods produce samples drawn from specific distributions by transforming a sequence of independent uniform deviates. An important reference on non-uniform random variate generation is Devroye (1986), many algorithms can also be found in Press et al. (2007).

2. NUMERICAL INTEGRATION

Before moving forward, it is interesting to mention the usefulness of estimating I based on sequences of *Common Random Numbers* (CRN), as in many applications I depends on parameters, say θ , that have to be estimated (see Hendry, 1984). The use of different set of random numbers for the evaluation of $\bar{I}_N(\theta)$ at various values of θ introduces an extra sort of variation in $\bar{I}_N(\theta)$ coming from the different random variates. This extra variation would imply excessive wiggling of $\bar{I}_N(\theta)$, rendering as problematic the application of numerical optimization algorithms. The use of CRN makes $\bar{I}_N(\theta)$ a smoother function of the parameters θ , facilitating the use of optimization procedures.

More specifically, the random numbers $\{x_i\}_{i=1}^N$ used in the evaluation of $\bar{I}_N(\theta)$ for all values of θ should be obtained as a non-stochastic transformation of a single set $\{u_i\}_{i=1}^N$ of canonical random draws, meaning that they are drawn from a distribution which is independent of any parameter of interest, like a uniform distribution on $(0, 1)$, or the standard normal distribution. The application of CRNs generates positive correlations between estimates of $\bar{I}_N(\theta)$ in neighboring values of θ , which is often enough to ensure sufficient smoothness for numerical optimization algorithms to succeed.

2.2.2 Importance Sampling

Although unbiased and straightforward, classical Monte Carlo described in the previous subsection is not the most used MC method as its variance can be too large. There are a variety of other more efficient techniques that are still unbiased, but can achieve variance reduction. Importance Sampling is a type of Monte Carlo method that tries to obtain variance reduction sampling $\{x_i\}_{i=1}^N$ not from the original density $f(x)$, but from a more suitable one. Note that the classical MC method estimates the integral (2.24) sampling from f in a manner completely unrelated to g , which is highly inefficient when larger values of g are concentrated in a relative small part of the domain of integration. The objective of importance sampling is to concentrate the distribution of the sample points in parts of the region of integration that are more important, instead of spreading them out irrespectively of g . To achieve this goal, importance sampling procedures modify the representation of the integral changing the probability measure from which sequences are generated. Hammersley and Handscomb (1964) popularized this

2.2 Monte Carlo Integration

method in their influential monograph. It was brought to the attention of econometricians by Kloek and van Dijk (1978) as a means to computing moments of posterior densities, and the conditions under which importance sampling approximations converge as well as the conditions under which numerical accuracy may be assessed reliably were set forth by Geweke (1989a).

Let $m(x)$ denote an auxiliary density function whose support contains that of f , and for which a pseudorandom number generator is available. Another representation of (2.24) is given by

$$I = E_m \left[\frac{g(x) \cdot f(x)}{m(x)} \right] = \int_{\mathbb{R}^n} g(x) \cdot \frac{f(x)}{m(x)} \cdot m(x) dx, \quad (2.31)$$

where $m(x)$ is called the importance sampling density, or *importance sampler*, and $f(x)$ may be called the *natural sampler*. An importance sampling estimate of (2.31) is obtained as

$$\bar{I}_N^{IS} = \frac{1}{N} \sum_{i=1}^N w(x_i) \cdot g(x_i), \quad (2.32)$$

where

$$w(x_i) = \frac{f(x_i)}{m(x_i)}, \quad (2.33)$$

are the *importance sampling weights*, and x_i are drawn from the auxiliary density $m(x)$. In contrast to an approximation based on simple average as in the classical MC approach, importance sampling uses weighted average in its approximations, where the weights are given by (2.33). These weights correct for differences between the natural sampler f , and the importance sampler m ; for example, if $m(x)$ has larger variance than $f(x)$, there will be too many x_i out in the tails of f , and too few around its mean, which would bias the result. Importance sampling corrects for this fact giving less weight to oversampled regions, and more weight to undersampled regions.

By the strong law of large numbers \bar{I}_N^{IS} converges almost surely to the true value of the integral in (2.31) for any choice of m whose support is a superset of the support of f , and $E_m \left[\bar{I}_N^{IS} \right] = I \forall N$ (see Robert and Casella, 2005).

In order to use the central limit theorem to assess the accuracy of importance sampling approximation and to guarantee that the rate of convergence to I is

2. NUMERICAL INTEGRATION

$N^{-\frac{1}{2}}$, the additional requirement that $\int g^2 \frac{f^2}{m} dx$ is absolutely convergent is necessary to ensure that the variance of the importance sampling estimate is finite:

$$\begin{aligned} \sigma_{IS}^2 &= \int g^2(x) \cdot w^2(x) \cdot m(x) dx - \left(\int g(x) \cdot w(x) \cdot m(x) dx \right)^2 \\ &= \int g^2(x) \cdot \frac{f^2(x)}{m(x)} dx - \left(\int g(x) \cdot f(x) dx \right)^2 \\ &= \int g^2(x) \cdot \frac{f^2(x)}{m(x)} dx - I^2, \end{aligned} \tag{2.34}$$

and hence

$$\int g^2(x) \cdot \frac{f^2(x)}{m(x)} dx < \infty, \tag{2.35}$$

is needed for the variance to be well defined. This requirement is clearly not met when the selected importance sampler m has lighter tails than f , in which case the importance sampling weights are going to be unbounded. Richard and Zhang (2007) offer a diagnostic measure that is adept at detecting this problem. The measure compares the MC sampling variances of the product $w(x)g(x)$ under draws from two alternative parameterizations of the importance sampler: the selected one, and another one that inflates the variance of the chosen sampler by a factor of 3 to 5. The idea behind this measure is to produce outliers far in the tails of the importance sampler, exactly where the importance weights are expected to explode, and to check their impact on the variance of the estimate. Koopman et al. (2009) provide another test for the validity of this hypothesis based on extreme value theory.

Equation (2.31) shows that an integral is not intrinsically associated with a given distribution function. Judd (1998) stresses this point by showing that importance sampling can be seen as an application of the change of variables rule. To see this connection, let $x(u) : [0, 1] \rightarrow \mathbb{R}$ be a monotone increasing function, and let $u(x)$ be its inverse, which implies that $u = u(x(u))$, and by the chain rule $1 = u'(x(u)) x'(u)$. The change of variables rule allows the following transformations of the integral being evaluated:

$$\begin{aligned} \int_{\mathbb{R}} g(x) \cdot f(x) dx &= \int_0^1 x'(u) \cdot g(x(u)) \cdot f(x(u)) du = \int_0^1 \frac{g(x(u)) \cdot f(x(u))}{u'(x(u))} du \\ &= E_U \left[\frac{g(x(u)) \cdot f(x(u))}{u'(x(u))} \right], \end{aligned} \tag{2.36}$$

2.2 Monte Carlo Integration

where $u \sim U[0, 1]$, and the identity $x'(u(x)) = 1/u'(x)$ was used. Note that this means that the probability density function of $x = x(u)$ is given by $m(x) = u'(x)$, and that any monotonic function $x(u) : [0, 1] \rightarrow \mathbb{R}$ can be used to transform $\int_{\mathbb{R}} g(x) f(x) dx$ into an integral over $[0, 1]$, which has the interpretation of an expectation with respect to the density of u .

Exactly as with change of variables, importance sampling is most useful when the selection of the map $x(u) : [0, 1] \rightarrow \mathbb{R}$ or, alternatively, the selection of $m(x) = u'(x)$, is done in a clever way. As the goal of importance sampling is to use Monte Carlo integration to evaluate the alternative expectation representation in (2.36), a suitable map $m(x)$ might be one that minimizes the sampling variance given in (2.34). The following first order condition can be used to find the optimal importance sampler m^* that minimizes the MC variance

$$0 = \frac{\delta}{\delta m(x)} \left[\int g(x)^2 \cdot \frac{f(x)^2}{m(x)} dx - I^2 + \lambda \cdot \int m(x) dx \right], \quad (2.37)$$

where λ is a Lagrange multiplier on the constraint that $m(x)$ must integrate to 1, and δ denotes the functional derivative. This first order condition gives $0 = -g^2 f^2 / m^2 + \lambda$, implying

$$m^*(x) = \frac{|g(x) \cdot f(x)|}{\sqrt{\lambda}} = \frac{|g(x) \cdot f(x)|}{\int |g(x) \cdot f(x)| dx}, \quad (2.38)$$

where λ was selected so that m^* integrates to one. Using m^* in (2.38), the importance sampling variance becomes

$$\sigma_{IS}^2 = \sigma_{IS}^{2*} = \left(\int |g(x) \cdot f(x)| dx \right)^2 - \left(\int g(x) \cdot f(x) dx \right)^2, \quad (2.39)$$

which is equal to zero⁷. Unfortunately, knowledge of $\int |g \cdot f| dx$ is needed to be able to sample from m^* , but if $\int |g \cdot f| dx$ is known, than numerical integration is not needed.

Although the result in (2.38) is not very useful in practice, it is still important to have in mind the idea of searching for an importance sampler that makes the MC variance as small as possible. Note that (2.34) will have a small sampling variance if $g(x)w(x)$ is as “constant” as possible. As stated above, it cannot be

⁷Note that if $g \cdot f$ has more than one sign in the region of integration (2.39) is not zero, but it is always possible to change the integrand by a known constant to make it all of one sign, since this changes the integral only by a known constant.

2. NUMERICAL INTEGRATION

made wholly constant, but (2.34) shows that $m(x)$ should try to mimic $g(x)f(x)$, so that the product $g(x)w(x)$ will not vary much.

The selection of the importance sampler $m(x)$ is the biggest challenge of importance sampling as m should closely mimic the integrand $g(x)f(x)$, whereas $g(x)f(x)$ is so complicated that it requires Monte Carlo integration. Later on Efficient Importance Sampling (EIS) (see Richard and Zhang, 2007), a procedure whose objective is to select a near optimal importance sampler $m(x)$ that mimics the integrand $g(x)f(x)$, will be presented.

In Bayesian estimation problems as well as in the update stage of nonlinear filtering, the density $f(x)$ is known only up to an integrating constant χ . Let $f^*(x)$ denote the known kernel of the density $f(x)$, such that $\chi = \int f^*(x) dx$. Importance sampling can still be applied using the fact that

$$\frac{1}{N} \sum_{i=1}^N \frac{g(x_i) \cdot f^*(x_i)}{m(x_i)} \xrightarrow{a.s.} \chi \cdot I \quad , \quad \frac{1}{N} \sum_{i=1}^N \frac{f^*(x_i)}{m(x_i)} \xrightarrow{a.s.} \chi,$$

which implies that an estimate of I can be computed as

$$\bar{I}_N^{IS*} = \frac{\sum_{i=1}^N g(x_i) \cdot \frac{f^*(x_i)}{m(x_i)}}{\sum_{i=1}^N \frac{f^*(x_i)}{m(x_i)}} \quad (2.40)$$

$$= \sum_{i=1}^N W(x_i) \cdot g(x_i), \quad (2.41)$$

where

$$W(x_i) = \frac{\frac{f^*(x_i)}{m(x_i)}}{\sum_{i=1}^N \frac{f^*(x_i)}{m(x_i)}}, \quad (2.42)$$

are the *normalized importance weights*. This estimator also converges to I , but unlike (2.32), (2.41) is biased for finite N . Doucet and Johansen (2009) give the asymptotic bias of \bar{I}_N^{IS*} :

$$\lim_{N \rightarrow \infty} N \left(\bar{I}_N^{IS*} - I \right) = - \int \frac{f^2(x)}{m(x)} \cdot (g(x) - I) dx. \quad (2.43)$$

The task of assessing the accuracy of \bar{I}_N^{IS*} as an approximation of I is complicated by the ratio of terms in (2.40) (see Geweke, 1989a). Now both $E_f[f/w] = \int_{\mathbb{R}^n} f^2/m dx$, and $\int g^2 w f dx$ are required to be absolutely convergent, which suffice

to show that

$$\begin{aligned}
 \bar{I}_N^{IS*} &\xrightarrow{a.s.} I, & \sqrt{N} \left(\bar{I}_N^{IS*} - I \right) &\xrightarrow{d} \mathcal{N}(0, \sigma^2), \\
 \sigma^2 &= E_f \left\{ [g(x) - I]^2 \cdot \frac{f^*(x)}{m(x)} \right\} = \chi^{-1} \int_{\mathbb{R}^n} \left\{ [g(x) - I]^2 \cdot \frac{f^*(x)}{m(x)} \cdot f(x) \right\} dx \\
 s^{*2} &= \frac{N \sum_{i=1}^N [g(x_i) - I]^2 \cdot \frac{f^*(x_i)}{m(x_i)}}{\left[\sum_{i=1}^N \frac{f^*(x_i)}{m(x_i)} \right]^2} \xrightarrow{a.s.} \sigma^2.
 \end{aligned} \tag{2.44}$$

Although biased, \bar{I}_N^{IS*} might perform better than (2.32) in squared error loss terms. To see this, note that the bias and the variance of (2.41) are $\mathcal{O}(1/N)$, which means that the mean-squared error is asymptotically dominated by the variance term (see Doucet and Johansen, 2009).

In what follows the GHK simulator is introduced, which is an importance sampler design to overcome problems faced by the frequency simulator, a classical Monte Carlo sampler for estimating normal probabilities of rectangular domains that can be used to estimate probit-type models (see Gourieroux and Monfort, 2002). Although the GHK does produce some variance reduction with respect to the frequency simulator, it is not aimed at achieving optimality in variance terms.

2.2.2.1 GHK importance sampler

An ubiquitous problem in probabilistic analysis and statistics is the calculation of event probabilities generated by random processes. Consider a n -dimensional vector of correlated normally distributed random variables: $x \sim N(\mu, \Sigma)$. Then, the probability that the event $A = \{(x_1, x_2, \dots, x_n) : a_i < x_i < b_i\}$ occurs is,

$$I \equiv \text{Prob}(A; \mu, \Sigma) = \int_{\mathbb{R}^n} \mathbb{I}(x \in D) \cdot f(x) dx, \tag{2.45}$$

where $\mathbb{I}(\cdot)$ is the indicator function, D is the rectangle defined by the limits a_i and b_i , and $f(x)$ is the probability density function of x .

If n is small, Gauss-Hermite quadrature could be used to compute (2.45), but it is often the case that n is large, and one has to rely on Monte Carlo estimates of this probability. Lerman and Manski (1981) present a simple Monte Carlo approximation to this integral obtained by drawing realizations of x from f , observing if a given draw lies within D , and using the frequency of such

2. NUMERICAL INTEGRATION

occurrences over the set of all draws as an estimate of the event probability I . This approach is known as the frequency simulator, and it has several drawbacks. First, it is not very efficient for approximating small probabilities, requiring a rather large number of draws, and even when the number of draws are moderately large, the approximation to I can be zero, which is problematic if one wants to compute the logarithm of this quantity. Second, the indicator function in (2.45) gives either weight equal to one or equal to zero to each draw from f , which means that the frequency simulator is a step function in the parameters $\theta = (\mu, \Sigma)$. For most values of θ , a small parameter change will change none of the weights given to the realizations of x , implying that the derivative of the frequency simulator in this direction will be zero. For the remaining values of θ , small changes in it will change many of the weights, and the simulator will increase (decrease) by a discrete amount as the weights changed from 0 to 1 (1 to 0). The derivative of the frequency simulator in this new direction will be undefined. As θ is usually estimated optimizing a criterion function based on (2.45), this lack of smoothness introduces numerical difficulties.

A very popular importance sampler that solves the aforementioned problems of the frequency simulator is the GHK (Geweke-Hajivassiliou-Keane) simulator introduced by Geweke (1989b), modified by Hajivassiliou et al. (1996), and independently proposed by Keane (1994) (see Train, 2003). The GHK is based on the fact that normal random variables conditional on other normal random variables are still normal, it exploits the following recursive structure imposed by a Cholesky decomposition

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} S_{1,1} & 0 & \cdots & 0 \\ S_{2,1} & S_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,1} & S_{n,2} & \cdots & S_{n,n} \end{bmatrix} \cdot \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}, \quad (2.46)$$

where the matrix S with elements $S_{i,j}$ is the Cholesky decomposition of Σ . Equation (2.2.2.1) implies that

$$I = \text{Prob}(A; \mu, \Sigma) = \int_{\mathbb{R}^n} \mathbb{I}(x(\epsilon) \in D) \cdot f(\epsilon) d\epsilon, \quad (2.47)$$

where $\epsilon \sim N(0, I_n)$, and I_n is a $(n \times n)$ identity matrix. Suppose that D is the

negative orthant of \mathbb{R}^n , then

$$\begin{aligned}
 x_1 < 0 &\rightarrow \epsilon_1 < -\frac{\mu_1}{S_{1,1}} \\
 x_2 < 0 &\rightarrow \epsilon_2 < -\frac{(S_{2,1}\epsilon_1 + \mu_2)}{S_{2,2}} \\
 &\vdots \\
 x_n < 0 &\rightarrow \epsilon_n < -\frac{(S_{n,1}\epsilon_1 + S_{n,2}\epsilon_2 + \dots + S_{n,n-1}\epsilon_{n-1} + \mu_n)}{S_{n,n}},
 \end{aligned} \tag{2.48}$$

The expressions in (2.48) can be written more compactly as

$$\begin{aligned}
 x_1 < 0 &\rightarrow \epsilon_1 < \alpha_1 \\
 x_2 < 0 &\rightarrow \epsilon_2 < \alpha_2(\epsilon_1) \\
 &\vdots \\
 x_n < 0 &\rightarrow \epsilon_n < \alpha_n(\epsilon_1, \dots, \epsilon_{n-1}),
 \end{aligned} \tag{2.49}$$

where

$$\alpha_n = \begin{cases} -\left(\sum_{i=1}^{n-1} \frac{S_{n,i}}{S_{n,n}} \epsilon_i + \frac{\mu_n}{S_{n,n}}\right), & n > 1 \\ -\frac{\mu_n}{S_{n,n}}, & n = 1. \end{cases} \tag{2.50}$$

Therefore, the event probability can be rewritten as

$$I = \text{Prob}(\epsilon_1 < \alpha_1, \epsilon_2 < \alpha_2(\epsilon_1), \dots, \epsilon_n < \alpha_n(\epsilon_1, \dots, \epsilon_{n-1})) \tag{2.51}$$

and the GHK simulator decomposes this joint probability of the sequence of events into a product of conditional event probabilities

$$I = \text{Prob}(\epsilon_1 < \alpha_1) \cdot \text{Prob}(\epsilon_2 < \alpha_2(\epsilon_1) | \epsilon_1) \cdot \dots \cdot \text{Prob}(\epsilon_n < \alpha_n(\epsilon_1, \dots, \epsilon_{n-1}) | \epsilon_1, \dots, \epsilon_{n-1}).$$

Because of the normality assumption on the errors $\{\epsilon_i\}_{i=1}^n$

$$\tilde{I}(\epsilon) = \Phi(\alpha_1) \cdot \Phi(\alpha_2(\epsilon_1)) \cdot \Phi(\alpha_3(\epsilon_2)) \cdot \dots \cdot \Phi(\alpha_n(\epsilon_1, \dots, \epsilon_{n-1})), \tag{2.52}$$

where Φ denotes the standard normal cumulative distribution function (cdf). This decomposition of the event probability motivates the use of a sequential importance sampler that draws ϵ_i from a truncated univariate normal distribution conditional on the last draw ϵ_{i-1} . A draw ϵ_1^s from the first element of the vector of errors is obtained from the univariate truncated normal pdf $\phi(\epsilon_1)/\Phi(\alpha_1)$, which is used to obtain a draw ϵ_2^s from $\phi(\epsilon_2)/\Phi(\alpha_2(\epsilon_1^s))$, which in turn is used to draw ϵ_3^s , and so on. The GHK importance sampler is therefore

$$m(\epsilon) = \begin{cases} \frac{\phi(\epsilon_1)}{\Phi(\alpha_1)} \cdot \frac{\phi(\epsilon_2)}{\Phi(\alpha_2(\epsilon_1))} \cdot \dots \cdot \frac{\phi(\epsilon_n)}{\Phi(\alpha_n(\epsilon_1, \dots, \epsilon_{n-1}))} & \text{for } \epsilon \in D, \\ 0 & \text{for } \epsilon \notin D. \end{cases} \tag{2.53}$$

2. NUMERICAL INTEGRATION

Note the fact that $f(\epsilon)$ in (2.47) is the numerator of $m(x)$ when $\epsilon \in D$, while $\tilde{I}(\epsilon)$ is the denominator, can be used to rewrite (2.53) as

$$m(\epsilon) = \begin{cases} \frac{f(\epsilon)}{\tilde{I}(\epsilon)} & \text{for } \epsilon \in D, \\ 0 & \text{for } \epsilon \notin D. \end{cases} \quad (2.54)$$

The GHK importance sampling representation of (2.45) is then given by

$$\begin{aligned} I &= \int_{\mathbb{R}^n} \mathbb{I}(x \in D) \cdot f(x) dx \\ &= \int_{\mathbb{R}^n} \mathbb{I}(x(\epsilon) \in D) \cdot f(\epsilon) d\epsilon \\ &= \int_{\mathbb{R}^n} \mathbb{I}(x(\epsilon) \in D) \cdot \frac{f(\epsilon)}{m(\epsilon)} \cdot m(\epsilon) d\epsilon \\ &= \int_{\mathbb{R}^n} \mathbb{I}(x(\epsilon) \in D) \cdot \frac{f(\epsilon)}{\frac{f(\epsilon)}{\tilde{I}(\epsilon)}} \cdot m(\epsilon) d\epsilon \\ &= \int_{\mathbb{R}^n} \mathbb{I}(x(\epsilon) \in D) \cdot \tilde{I}(\epsilon) \cdot m(\epsilon) d\epsilon \\ &= \int_{\mathbb{R}^n} \tilde{I}(\epsilon) \cdot m(\epsilon) d\epsilon, \end{aligned} \quad (2.55)$$

where the last equality follows from the fact that $m(\epsilon) > 0$ only when $\epsilon \in D$. An estimate of (2.55) can be obtained as

$$\bar{I}_N = \frac{1}{N} \sum_{s=1}^N \Phi(\alpha_1) \cdot \Phi(\alpha_2(\epsilon_1^s)) \cdot \Phi(\alpha_3(\epsilon_2^s)) \cdot \dots \cdot \Phi(\alpha_n(\epsilon_1^s, \dots, \epsilon_{n-1}^s)), \quad (2.56)$$

where the ϵ_i^s are drawn from the truncated normal densities in (2.53) using $\Phi^{-1}(u \cdot \Phi(\alpha_n(\cdot)))$, with u being a random number taken from $U_{(0,1)}$.

The GHK simulator replaces the discrete weighting scheme of the frequency simulator for a continuous one via a clever choice of importance sampler. The estimate (2.56) is strictly bounded between $(0, 1)$, continuous, and differentiable with respect to μ and Σ , because $\tilde{I}(\epsilon)$ is also strictly bounded between $(0, 1)$, continuous, and differentiable. Although the GHK simulator does not seek directly variance reduction, its variance is smaller than the variance of the frequency simulator, because each draw from $m(\epsilon)$ receives weight between $(0, 1)$, while draws from the frequency simulator receive either weight equal to 1 or 0. However, it is interesting to note that the GHK importance sampler is not a good

simulator of draws from the joint density of x . The GHK samples x_i and incorporates the constraints that $(x_1, \dots, x_i) < 0$, but ignores the correlated information $(x_{i+1}, \dots, x_n) < 0$, neglecting potentially significant information that would allow to better adjust the region of importance for x_i . Based on this, Liesenfeld and Richard (2007) argue that the GHK importance sampler can be interpreted as a filtering density, incorporating the constraints on x just up to the element i , and efficiency gains could be sought by including the additional information contained in the other constraints.

2.2.3 Efficient Importance Sampling

The main drawbacks of IS are the difficulties in finding a good importance sampler (van Dijk, 1999). As explained above, a proper metric to define “good” in the context of Monte Carlo methods is variance reduction. Therefore, the choice of a good importance sampler could be divided into two steps: the selection of a parametric class of samplers; and the selection of the parameters of the chosen class that delivers an importance sampler that makes the sampling variance as small as possible.

The choice of a parametric class of samplers is problem specific and will vary from application to application, but the chosen class would normally include a parametric extensions of the natural sampler $f(x)$. In this subsection, the Efficient Importance Sampling (EIS) technique to find a near optimal sampler within a given parametric class is going to be presented. EIS was developed by Richard and Zhang (2007), and is an automated procedure for constructing importance samplers that are global approximations to the target integrand $g(x)f(x)$ in (2.31), in contrast to local approximations based on Taylor series expansions used in the literature (see, eg., Shephard and Pitt, 1997).

EIS is a generic auxiliary algorithm that searches in the admissible parameter space A of a given class $M : \{m(x; a); a \in A\}$ of parametric samplers for the optimal parameters a^* that determines the most efficient sampler within this class. The objective of EIS is the selection of an optimal parameter value a that minimizes the MC sampling variance of the integral (2.31). The MC sampling variance of a given importance sampler in (2.34) can be rewritten as function of

2. NUMERICAL INTEGRATION

the parameter a as:

$$\begin{aligned}\sigma_{IS}^2 &= \int \left[\frac{g(x) \cdot f(x)}{m(x; a)} - I \right]^2 \cdot m(x; a) dx \\ &= \int [\omega(x; a) - I]^2 \cdot m(x; a) dx,\end{aligned}\tag{2.57}$$

where $\omega(x; a) = \frac{g(x)f(x)}{m(x;a)}$. Rearranging to isolate the terms depending on a yields

$$\begin{aligned}\sigma_{IS}^2 &= I \cdot \int \left[\frac{\omega(x; a)}{I} - 1 \right]^2 \cdot \frac{I}{\omega(x; a)} \cdot g(x) \cdot f(x) dx \\ &= I \cdot \int \left[\frac{\omega(x; a)^2}{I^2} - 2\frac{\omega(x; a)}{I} + 1 \right] \cdot \frac{I}{\omega(x; a)} \cdot g(x) \cdot f(x) dx \\ &= I \cdot \int \left[\frac{\omega(x; a)}{I} + \frac{I}{\omega(x; a)} - 2 \right] \cdot g(x) \cdot f(x) dx \\ &= I \cdot V(x; a),\end{aligned}\tag{2.58}$$

where

$$V(x; a) = \int \left[\frac{\omega(x; a)}{I} + \frac{I}{\omega(x; a)} - 2 \right] \cdot g(x) \cdot f(x) dx,$$

is the part of σ_{IS}^2 depending on a . Let

$$d(x; a, c) = \ln \left(\frac{\omega(x; a)}{I} \right) = \ln (g(x) \cdot f(x)) - c - \ln k(x; a),\tag{2.59}$$

where the the importance sampler $m(x; a)$ was decomposed into a density kernel $k(x; a)$ and its integrating constant $\chi(a) = \int k(x; a) dx$ such that

$$m(x; a) = \frac{k(x; a)}{\int k(x; a) dx},\tag{2.60}$$

and thus

$$c = \ln I - \ln \chi(a),$$

is a constant meant to calibrate the ratio $g(x)f(x)/k(x; a)$ which does not depend on x but on the unknown quantity I , and thus will be treated as an unknown intercept in the optimization problem to be introduced. Now it is possible to write $V(x; a)$ as

$$V(x; a) = \int h(d^2(x; a, c)) \cdot g(x) \cdot f(x) dx,\tag{2.61}$$

with h given by

$$h(r) = e^{\sqrt{r}} + e^{-\sqrt{r}} - 2 = 2 \sum_{i=1}^{\infty} \frac{r^i}{(2i)!}.$$

An optimal choice of (a, c) can be found through the solution of the following nonlinear minimization problem

$$(a^*, c^*) = \arg \min_{a \in A, c \in \mathbb{R}} V(x; a). \quad (2.62)$$

Note that $h(r)$ is a monotone and convex function on \mathbb{R}_+ , and an efficient sampler will be one such that $m(x; a)$ closely mimics the integrand $g(x)f(x)$, implying that $d(x; a)$ is expected to be close to zero. Therefore, an useful simplification is attained by replacing $h(r)$ by its leading term r , which means solving the simpler problem

$$(\hat{a}, \hat{c}) = \arg \min_{a \in A, c \in \mathbb{R}} Q(x; a), \quad (2.63)$$

where

$$Q(x; a) = \int d^2(x; a) \cdot g(x) \cdot f(x) dx. \quad (2.64)$$

Equation (2.63) can be seen as a functional Generalized Least Squares (GLS) problem, where x is distributed according to the natural sampler $f(x)$ and weight $g(x)$. However, MC approximations of Q based on draws sampled from f would be generally inaccurate due to high MC sampling variance of g , and that is actually the reason why importance sampling is being used in the first place. Approximations of Q based on an efficient sampler $m(x; a)$ would be preferred, thus Q can be rewritten as

$$Q(x; a) = \int d^2(x; a) \cdot \omega(x; a) \cdot m(x; a) dx. \quad (2.65)$$

The EIS approach is to replace $Q(x; a)$ with a Monte Carlo approximation thereof, but the fact that the importance sampler itself depends on a implies that finding a direct solution to the minimization problem in (2.63) is not feasible. Instead, Richard and Zhang (2007) adopt an iterative procedure to search for a fixed point, whereby a sequence of samplers $\{m(x; \hat{a}^k)\}_{k=0}^*$ is constructed that converges towards $m(x; \hat{a}^*)$. Specifically, the baseline EIS algorithm consists of

2. NUMERICAL INTEGRATION

computing a converging sequence of GLS estimates of $\{\hat{a}^k\}_{k=0}^*$, which are used to construct global approximations to the target integrand, as they are based on GLS estimates on the full support of $g(x)f(x)$. The algorithm is based upon the following recursion:

$$(\hat{a}^{k+1}, \hat{c}^{k+1}) = \arg \min_{a \in A, c \in \mathbb{R}} \bar{Q}(x; a | \hat{a}^k), \quad (2.66)$$

where

$$\bar{Q}(x; a | \hat{a}^k) = \frac{1}{N} \sum_{i=1}^N d^2(x_i^k, a, c | \hat{a}^k, \hat{c}^k) \cdot \omega(x_i^k; \hat{a}^k), \quad (2.67)$$

and $\{x_i^k\}_{i=1}^N$ are i.i.d. draws from $m(x; \hat{a}^k)$. An initial sampler $m(x; \hat{a}^0)$ can be found in a variety of ways, as it will be shown in subsequent chapters, but a convenient choice for an initial sampler (although not the most efficient) is to take the natural sampler $f(x)$ as $m(x; \hat{a}^0)$. In this case, Richard and Zhang (2007) advise to set the weights ω to ones during the initial iteration(s) to avoid numerical instabilities due to the large variance of ω under an inefficient sampler like $f(x)$, and state that for most problems the Ordinary Least Squares (OLS) problem, where all the weights are set to one, is essentially as efficient as its GLS counterpart. A maximum relative change in a of order 10^{-3} to 10^{-5} is normally used as a convergence criterion, and a critical point to convergence is the use CRNs in the EIS auxiliary regressions in (2.66). Once convergence is achieved, the EIS estimate of (2.31) is given by

$$\bar{I}_N = \frac{1}{N} \sum_{i=1}^N \omega(x_i^*; \hat{a}^*). \quad (2.68)$$

2.2.3.1 EIS for the exponential family of distributions

The exponential family of distribution has densities of the form

$$m(x; a) = \chi(a) \cdot b(x) \cdot \exp \{a' \cdot t(x)\}, \quad (2.69)$$

where the set $A \supset a$ is called the natural parameter space of the family, a is a vector of natural parameters, and $t(x)$ is a sufficient statistic of fixed dimension for all sample sizes (see Lehmann and Romano, 2005; Lehmann and Casella,

1998). A prominent member of the exponential family is the normal distribution function, and if $X \sim N(\mu, \sigma^2)$, then

$$m(x; a) = \frac{1}{\sqrt{\frac{\pi}{-a_2}}} \cdot \exp \left[a_1 x + a_2 x^2 - \frac{a_1^2}{4a_2} \right],$$

is a two-parameter exponential family member with natural parameters $(a_1, a_2) = (\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})$, and natural parameter space $A = \mathbb{R} \times (-\infty, 0)$.

The logkernels of this family of distribution can be written as

$$\ln k(x; a) = \ln b(x) + a' \cdot t(x), \tag{2.70}$$

and direct substitution in the expression of $d(x; a, c)$ defined in (2.59) shows that the EIS auxiliary regressions are now linear in a . Assuming that the GLS weights $\omega(x; a)$ are nearly constant, it is possible to substitute the GLS regressions for standard OLS regressions, simplifying even more the EIS minimization problem.

The natural parameter space A typically imposes inequality constraints for the integrability of k ; for example, in the univariate normal distribution case, $a_2 \in (-\infty, 0)$. Although this constraints are normally not binding, in which case OLS or GLS formulae provide analytical solution to the minimization problem in (2.63), problems might occur when the initial sampler $m(x; a^0)$ concentrate draws in regions where $g(x)f(x)$ is locally convex, like the tail area or the area between two local peaks. Inflating the variance of the initial sampler and/or setting the corresponding coefficient a_i to some arbitrary value close to the boundary of the natural parameter space is a numerical way to solve such occasional pathologies.

Another important property of the exponential family is the convexity of the natural parameter space A , meaning that if $a^0 \in A$ and $a^1 \in A$ then $a^2 = a^0 + a^1$ also belongs to A . This implies that $K = \{k(x; a); a \in A\}$ is closed under multiplication (Lehmann and Casella, 1998), resulting in possible additional simplifications of the EIS auxiliary regression when there exists $a^0 \in A$ such that

$$g(x)f(x) \propto g_0(x) \cdot k(x; a^0), \tag{2.71}$$

where $k(x; a^0)$ could denote $f(x)$ itself when $f \in M$, but could also include additional factors from g , improving the fit of $m(x; a)$ to $g(x)f(x)$, and $g_0(x) = g(x)$ in the case that $k(x; a^0) = f(x)$, or it contains the remaining terms of $g(x)$

2. NUMERICAL INTEGRATION

that could not be included in $k(x; a^0)$ ⁸. In this case it is possible to exploit the closure of K by redefining $k(x; a)$ as

$$k(x; a^2) = k(x; a^1) \cdot k(x; a^0),$$

where $a^2 = a^1 + a^0$ and following (2.59) $k(x; a^0)$ cancels out in the EIS Least Squares (LS) regression. The auxiliary regression then becomes a LS regression of $\ln g_0$ on $k(x; a^1)$.

To better understand the implementation of the EIS procedure for Gaussian importance samplers, note that the idea in (2.59) is to approximate $g(x)f(x)$ by a Gaussian kernel

$$\begin{aligned} \ln[g(x)f(x)] &\approx -\frac{1}{2}(x - \mu)'H(x - \mu) \\ &\approx -\frac{1}{2}(x'Hx - 2x'H\mu), \end{aligned}$$

where μ is the mean vector and H the precision matrix associated with the natural parameter a . The term $x'Hx$ can be written as

$$\begin{aligned} &\begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,n} \\ h_{2,1} & h_{2,2} & \dots & h_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n,1} & h_{n,2} & \dots & h_{n,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &= h_{1,1}x_1^2 + h_{2,2}x_2^2 + \dots + h_{n,n}x_n^2 \\ &\quad + 2h_{1,2}(x_1x_2) + 2h_{1,3}(x_1x_3) + \dots + 2h_{1,n}(x_1x_n) \\ &\quad + 2h_{2,3}(x_2x_3) + 2h_{2,4}(x_2x_4) + \dots + 2h_{2,n}(x_2x_n) \\ &\quad \vdots \\ &\quad + 2h_{n-1,n}(x_{n-1}x_n). \end{aligned}$$

The relationship between the coefficient of the squared and pairwise product of the elements of x , and the components of the precision matrix H , as well as the relationship between the individual components of x and the vector of means μ are evident from the above expressions. These coefficients are the natural parameters

⁸Note that $g(x)f(x) \propto k(x; a^0)$ implies a perfect fit, meaning that $k(x; a^0)$ include all factors of g depending on x .

of the normal distribution, and EIS tries to estimate the natural parametrization that makes the importance sampler as close to the target integrand as possible.

As a concrete example of an univariate auxiliary EIS regressions for the Gaussian class of densities, consider the linear regression in the k^{th} iteration of the EIS fixed point search (2.66) for a Gaussian univariate sampler

$$\ln[g(x_i) \cdot f(x_i)] = c + a_1^k x_i + a_2^k x_i^2 + u_i, \quad (2.72)$$

where $x = \{x_i\}_{i=1}^N$ were drawn from the sampler $m(x; \hat{a}^{k-1})$ of the last EIS iteration, and u_i is the error term. The new LS estimates of a are used to construct the updated sampler $m(x; \hat{a}^k)$ as

$$m(x; \hat{a}^k) = \frac{1}{\sqrt{\frac{\pi}{-\hat{a}_2^k}}} \cdot \exp \left[\hat{a}_1^k x + \hat{a}_2^k x^2 - \frac{(\hat{a}_1^k)^2}{4\hat{a}_2^k} \right],$$

with mean and variance given by

$$\sigma^2(\hat{a}^k) = -\frac{1}{2\hat{a}_2^k}, \quad \mu(\hat{a}^k) = -\frac{\hat{a}_1^k}{2\hat{a}_2^k}. \quad (2.73)$$

When the dimensionality of a is large with respect to the MC sample $\{x_i\}_i^N$ used in the auxiliary regression (2.72), problems of multicollinearity can occur. In such cases, introduction of a small amount of shrinkage in the auxiliary EIS regressions is able to eliminated the problem, and is numerically more reliable than increasing N (see Richard and Zhang, 2007).

2.2.3.2 Sequential EIS

Although shrinkage EIS regressions can alleviate multicollinearity problems arising when the dimensionality of the objective integral (and thus that of a) is large, in high-dimensional settings the problem (2.63) needs to be replaced by a sequence of lower dimensional optimization problems. This factorization is possible when the problem at hand can be written not only as a single joint distribution, but also as a sequence of conditional distribution. This is often the case in the time series dimension, where the latent and observable variables are typically generated sequentially in time, but can also happen in the cross-sectional dimension when correlation across units exists⁹.

⁹The GHK importance sampler presented in section 2.2.2.1 exploited such a decomposition to construct a sequential importance sampler

2. NUMERICAL INTEGRATION

Assume that a partition $x = (x_1, x_2, \dots, x_L)$ of x in lower dimension components exists, allowing to decompose g and f in (2.31) as

$$\begin{aligned} g(x) &= \prod_{l=1}^L g_l(X_l) \\ f(x) &= \prod_{l=1}^L f_l(x_l|X_{l-1}) \\ I &= \int g(x) \cdot f(x) dx = \int \prod_{l=1}^L g_l(X_l) \cdot f_l(x_l|X_{l-1}) dx, \end{aligned} \quad (2.74)$$

where $X_l = (x_1, \dots, x_l)$. In such cases, the EIS sampler $m(x; a)$ is subject to the following sequential structure

$$m(x; a) = \prod_{l=1}^L m_l(x_l|X_{l-1}; a_l), \quad (2.75)$$

where $a = (a_1, \dots, a_L) \in A = \prod_{l=1}^L A_l$, as well as its kernel k_l and integrating constant χ_l

$$m_l(x_l|X_{l-1}; a_l) = \frac{k_l(X_l; a_l)}{\chi_l(X_{l-1}; a_l)}, \quad \chi_l(X_{l-1}; a_l) = \int k_l(X_l; a_l) dx_l.$$

However, the integral of $g_l(X_l)f_l(x_l|X_{l-1})$ with respect to x_l is an unknown function of X_{l-1} and cannot be well approximated by $m_l(x_l|X_{l-1}; a)$, whose integral with respect to x_l is by definition equal to one. Instead, the l^{th} integrand $g_l(X_l)f_l(x_l|X_{l-1})$ is approximated by a density kernel $k_l(X_l; a_l)$ with known functional integral $\chi_l(X_{l-1}; a_l)$ in x_l , so that once \hat{a}_l is selected, an analytical expression for χ_l is available. This would leave the integrating constant $\chi_l(X_{l-1}; a_l)$ unaccounted for in the EIS optimization procedure, disregarding possibly important information about X_{l-1} contained in $\chi_l(X_{l-1}; a_l)$ ¹⁰. Since $\chi_l(X_{l-1}; a_l)$ does not depend on X_l , but only on X_{l-1} it could be transferred back to the LS optimization on a_{l-1} , so that the efficient importance sampler $m_l(x_l|X_{l-1}; \hat{a}_l)$ will contain all the sample information about the importance region of x_l . The integral (2.74) is then rewritten as

$$I = \chi_1(a_1) \int \prod_{l=1}^L \frac{g_l(X_l)f_l(x_l|X_{l-1})\chi_{l+1}(X_l; a_{l+1})}{k_l(x_l; a_l)} \cdot \prod_{l=1}^L m_l(x_l|X_{l-1}; a_l) dx, \quad (2.76)$$

¹⁰Note that this is exactly the information disregarded by the GHK importance sampler.

2.2 Monte Carlo Integration

with $\chi_{L+1} \equiv 1$. In order to allow the back transfer of all relevant integrating constants to the LS optimization of a_l , a backward recursive sequence of individual EIS minimizations starting from $l = L$ and going until $l = 1$ is implemented as follows

$$(\hat{a}_l^{k+1}, \hat{c}_l^{k+1}) = \arg \min_{a_l \in A_l, c_l \in \mathbb{R}} \bar{Q}(X_l; a_l, \hat{a}_{l+1}), \quad (2.77)$$

where

$$\bar{Q}(X_l; a_l, \hat{a}_{l+1}) = \frac{1}{N} \sum_{i=1}^N d_l^2(X_{i,l}; a_l, \hat{a}_{l+1} c_l) \cdot \omega_l(X_{i,l} | X_{i,l-1}; a_l),$$

and

$$d_l(X_{i,l}; a_l, \hat{a}_{l+1}, c_l) = \{\ln [g_l(X_{i,l}) f_l(x_{i,l} | X_{i,l-1}) \chi_{l+1}(X_{i,l}; \hat{a}_{l+1})] - c - \ln k_l(x_{i,l}; a_l)\}$$

now takes into account the integrating constant $\chi_{l+1}(X_{i,l}; \hat{a}_{l+1})$ of the previous period, which contains possibly important information with respect to the importance region of x_l .

Once the backward sequence of EIS regressions has converged to a fixed point \hat{a}^* , a MC estimate of (2.31) is obtained as

$$\bar{I}_N = \frac{1}{N} \sum_{i=1}^N \left[\prod_{l=1}^L \frac{g_l(X_{i,l}) f_l(x_{i,l} | X_{i,l-1})}{m_l(x_{i,l} | X_{i,l-1}; \hat{a}_l)} \right], \quad (2.78)$$

where the sequences $\{X_{i,l}\}_{i=1}^N$ are drawn from the efficient importance sampler $m(x; \hat{a}^*)$.

The decomposition in (2.76) and the backward algorithm in (2.77) show that the EIS sampler $m_l(x_l | X_{l-1}; a_l)$ contains all sample information about the importance region of x_l , in sharp contrast to the GHK importance sampler presented in section 2.2.2.1. In this sense, the EIS sampler could be regarded as smoothed densities, while the GHK densities are filtered densities as they consider only information up to the observation l .

2. NUMERICAL INTEGRATION

Chapter 3

Estimation of Dynamic Panel Probit Models

Estimation of panel probit applications incorporating unobserved factors requires solving integrals over latent variables whose solution are not analytically available, implying that numerical integration methods have to be used. Typically these integrals are multidimensional restricting the use of deterministic quadrature methods to the simplest and less interesting models. In this chapter different panel probit models accounting for unobserved heterogeneity and serially correlated errors are estimated and used to analyze the determinants and the dynamics of current-account reversals for a panel of developing and emerging countries.

Likelihood evaluation of panel probit models with unobserved heterogeneity and dynamic error components is complicated by the fact that the computation of the choice probabilities requires high-dimensional interdependent integrations. The dimension of such integrals is typically given by the number of time periods (T), or if one allows for interaction between individual specific and time random effects by $T+N$, where N is the number of cross-sectional units. Thus likelihood estimation of such models generally relies upon Monte Carlo integration techniques (see Geweke and Keane, 2001, and the references therein). Various MC procedures have been proposed for the evaluation of such event probabilities (see Stern, 1997, for a survey). The most popular among those is the GHK procedure presented in Section 2.2.2.1, which has been applied to the estimation of dynamic panel probit models, e.g., by Hyslop (1999), Greene (2004), and Falchetti and Tudela (2006). While conceptually simple and easy to program, the

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

GHK procedure relies upon importance sampling densities which ignore critical information relative to the underlying dynamic structure of the model. This can lead to significant deterioration of numerical accuracy as the dimensionality of integration increases. In particular, Lee (1997) conducts a MC study of ML estimation under GHK likelihood evaluation for panel models with serially correlated errors, and finds significant biases for longer panels. The Efficient Importance Sampling (EIS) methodology developed by Richard and Zhang (2007) presented in Section 2.2.3 is used here to evaluate the event probabilities, and is shown to be well suited to handle unobserved heterogeneity and serially correlated errors in panel probit models. In particular, combining EIS with GHK substantially improves the numerical efficiency of the standard GHK allowing for reliable maximum likelihood estimation of dynamic panel probit models even in applications with a large time dimension.

3.1 Introduction

The determinants of current account reversals and their consequences for countries' economic performance have received a lot of attention since the currency crises of the 1990s, and have found renewed interest because of the huge current account deficit of the US in recent years. The importance of the current account comes from its interpretation as a restriction on countries' expenditure abilities. Expenditure restrictions, generated by sudden stops and/or currency crises, can generate current account reversals, worsen an economic crises or even trigger one (see Milesi-Ferretti and Razin, 1997, 1998, 2000; Obstfeld and Rogoff, 2007). Typical issues addressed in the recent literature are: the extent to which current account reversals affect economic growth (Edwards, 2004a,b; Milesi-Ferretti and Razin, 2000); the sustainability of large current account deficits for significant periods of time (Milesi-Ferretti and Razin, 2000); and possible causes for current account reversals (Edwards, 2004a,b; Milesi-Ferretti and Razin, 1998). This chapter proposes to analyze the latter issue in the context of dynamic panel probit models, paying special attention to the serial dependence inherent to current account reversals.

Milesi-Ferretti and Razin (1998) and Edwards (2004b) use panel probit models in order to investigate the determinants of current account reversals. While

Milesi-Ferretti and Razin analyze a panel of low- and middle-income countries, Edwards also includes industrialized countries. They use time and country specific dummies in order to account for heterogeneity. In addition to the fact that it requires estimation of a large number of parameters, a fixed effect approach raises two key issues of identification in the context of the data set analyzed here. First, it precludes the use of potentially important explanatory variables which are constant across countries or over time. Also, current account crises are typically rare events and have not been experienced by some of the countries included in our data set.

Those studies paid less attention to the potential intertemporal linkages among current account reversals, focussing on tests of theoretical predictions about their causes, which are mainly motivated by the need to ensure that a country remains solvent. However, there are several reasons to expect serial persistence in current account reversals. For example, a full current account adjustment from a non-sustainable towards a sustainable level might take several periods since responses of international trade flows are characterized by a fairly high degree of inertia (see Junz and Rhomberg, 1973). Furthermore, past current account reversals might change the constraints and conditions relevant to the occurrence of another reversal in the future, as argued by Falcetti and Tudela (2006) within the context of a panel analysis of currency crisis. Both scenarios would lead to *state dependence* (lagged dependent variable), whereby a country's propensity to experience a reversal depends on whether or not it experienced a reversal in the past (see Heckman, 1981). Following Falcetti and Tudela (2006), further potential sources of serial dependence are unobserved time-invariant heterogeneity (random country specific effect) reflecting differences in institutional, political or relevant economic factors across countries, as well as unobserved transitory differences (country specific serially correlated error term) which might be the result of omitted serially correlated macroeconomic factors or serially correlated country-specific shocks.

However, unobserved and serially correlated transitory effects might be also common to all countries (common dynamic time random effect). As such they might reflect global shocks like oil and other commodity price shocks affecting the reversal probability of all countries, or might be the result of contagion effects. In particular, following the financial turbulence of the 1990s, it is recognized

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

that spillover effects are important, especially for emerging economies. Common causes of contagion include transmission of local shocks, such as currency crises, through trade links, competitive devaluations, and financial links (see Dornbusch et al., 2000).

3.2 Determinants and Dynamics of Current Account Reversals

3.2.1 Determinants

Milesi-Ferretti and Razin (1996) argue that the most obvious reason for a country to experience a current account reversal is the need to ensure solvency, which they relate to the stabilization of the ratio of external liabilities to GDP. Let tb be the trade balance before the reversal and tb^* the trade balance needed to stabilize the ratio of external liabilities to GDP. Then abstracting from equity and foreign direct investment flows and stocks, the reversal needed to ensure solvency can be written as:

$$\begin{aligned} REV &= tb^* - tb = (r^* - \epsilon^* - \gamma^*)d - tb \\ &[(r^* - r) - \epsilon^* - \gamma^*]d - (s - i), \end{aligned} \tag{3.1}$$

where r is the real interest rate on external debt, γ is the growth rate of the economy, ϵ is the rate of real appreciation, d is the ratio of external debt to GDP, s is the share of domestic savings to GDP, and i is the share of domestic investment to GDP. Variables indexed by a star denotes the level of variable x needed to stabilize the external debt to GDP ratio.

This simple framework points to several determinants for the occurrence of large reductions in the current-account imbalance. The size of the reversal needed to ensure solvency grows with the initial trade imbalance. Given the initial trade imbalance, the size of the required reversal is increasing in the level of external liabilities as well as in the rate of interest on external debt, while it is decreasing in the growth rate. Note also that an increase in the world interest rate lowers the interest rate differential, increasing the required reversal size. In fact, any change in r^* and γ^* will affect a country's intertemporal budget constraint, and thus its current account imbalance.

3.2 Determinants and Dynamics of Current Account Reversals

Further potential determinants for current account reversals are obtained from models developed to analyze the ability of a country to sustain a large current account deficit for significant periods of time (see Milesi-Ferretti and Razin, 1998). They indicate that the sustainability of an external imbalance and, therefore, the probability of its reduction depend on factors such as a country's degree of openness, its international reserves, its terms of trade and its fiscal environment.

While the solvency condition (3.1) helps identifying potential causes for the occurrence of current account reversals, it is static and, therefore, not helpful to discuss the dynamics of reversals. However, as discussed further below, there are several reasons to expect serial dependence in the occurrence of large reductions of current account deficits. Within a panel probit model for the analysis of the determinants of reversals they imply state dependence and/or serially correlated error terms.

3.2.1.1 State dependence

Assuming that the domestic economy for a deficit country ($d > 0$) grows at a rate below the real interest rate (adjusted by the rate of real appreciation), the solvency condition (3.1) requires a trade surplus stabilizing the debt to GDP ratio. This surplus is often obtained by currency devaluations. However, while changes in exchange rate can be abrupt, subsequent changes in trade can be much slower. Junz and Rhomberg (1973) analyze the response of international trade flows to changes in the exchange rate, and conclude that the effects of price changes on trade flows usually stretch out over more than three years. In particular, they argue that agents react with lags and identify the following sources for delayed responses: a recognition lag, which is the time it takes for economic agents to become aware of changes in the competitive environment; a decision lag, which lasts from the moment in which the new situation has been recognized to the one in which an action is undertaken (producers need to be convinced that the new opportunities are long lasting and profitable enough to compensate for adjustment costs); and finally, mostly technical lags in production, delivery and substitution of materials and equipments in response to relative price changes.

In line with these arguments, Himarios (1989) finds that nominal devaluations result in significant real devaluations that last for at least three years, and that real devaluations induce significant trade flows that are distributed over a two- to

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

three-year period. Therefore, the full current account adjustment implied by (3.1) might take longer than one year, leading to a state dependence for yearly data such as those used below. In order to account for the possibility that a reversal process stretches over more than a year after it is triggered, the lagged dependent variable is included among the regressors of our panel probit specifications.

3.2.1.2 Serially correlated error terms

Further potential sources of serial dependence in the occurrence of large reductions in the current account imbalance are differences in the propensity to experience large reductions across countries. Such heterogeneity might be due to time-invariant differences in institutional, political or economic factors which can not be controlled for. In order to take these differences into account, a random effect approach with a country-specific time-invariant error component is used, which induces a cross-period correlation of the overall error terms. Unobserved differences in the propensity to experience large reductions in the current account deficit could also be serially correlated, rather than time-invariant. As such they might reflect serially correlated shocks associated with regional conflicts, uncertainty about government transition and political changes, as well as regional commodity price shocks affecting the probability of experiencing current account reversals. In order to take those effects into account, an AR(1) specification for the country specific transitory error component is assumed.

Finally, unobserved and serially correlated transitory effects might also be common to all countries reflecting either contagion effects or global shocks such as oil or commodity price shocks. The former have received a lot of attention following the currency crises of the 1990s which rapidly spread across emerging countries (see Edwards and Rigobon, 2002). A crisis in one country may affect other countries as investors withdraw their investments from other markets without taking into account differences in economic fundamentals. Contagion can also happen through trade links and currency devaluations, as devaluations foment output growth and employment domestically at the expense of output growth, employment and current account deficit abroad, which could even incite a process of *competitive devaluations* (Corsetti et al., 1999).

If data are collected at short enough time intervals (monthly or quarterly observations), such spill-over effects would become manifest in the dependence

3.2 Determinants and Dynamics of Current Account Reversals

of a country's propensity to experience a reversal from lagged reversals by other countries. However, with yearly data the time intervals are presumably not fine enough to observe such short-run spill-over effects of one country on another and contagion would more likely translate into a common time effect. Hence, an AR(1) time-random effect which is common to all countries is used in order to account for contagion effects together with global shocks.

3.2.2 Data

The data set consists of an unbalanced panel for 60 low and middle income countries from Africa, Asia, and Latin America and the Caribbean. The complete list of countries is given on Table 3.7. The time span of the data set ranges from 1975 to 2004, although the unavailability of some explanatory variables often restrict the analysis to shorter time intervals. The minimum number of periods for a country is 9, the maximum is 18 and the average is 16.5 for a total of 963 yearly observations. The initial values of the binary dependent variable indicating the occurrence of a current account crisis are known for the initial time period $t = 0$ for all countries. The sources of the data are the World Bank's World Development Indicators (2005) and the Global Development Finance (2004).

Current account reversals are defined as in Milesi-Ferretti and Razin (1998).

According to this definition a current account reversal in year t has to meet three requirements. The first is an average reduction of the current account deficit of at least 3 percentage points of GDP over a period of 3 years (from t to $t + 2$) relative to the 3-year average before the event (from $t - 3$ to $t - 1$). The second requirement is that the maximum deficit in the 3-year period starting with the reversal must be no larger than the minimum deficit in the 3 years preceding the reversal. The last requirement is that the average current account deficit over the 3-year period starting with the event must be less than 10% of GDP. According to this definition, 100 current account reversals are found in 44 countries (10% of the total number of observations).

As discussed above, the selection of the explanatory variables follows mainly the study of Milesi-Ferretti and Razin (1998). Lagged macroeconomic, external, debt and foreign variables are included. The macroeconomic variables are the annual growth rate of GDP (AVGGROW), the share of investment to GDP proxied by the ratio of gross capital formation to GDP (AVGINV), government

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

expenditure (GOV) and interest payments relative to GDP (INTPAY). The external variables are the current account balance as a fraction of GDP (AVGCA), a terms of trade index set equal to 100 for the year 2000 (AVGTT), the share of exports and imports of goods and services to GDP as a measure of trade openness (OPEN), the rate of official transfers to GDP (OT) and the share of foreign exchange reserves to imports (RES). The debt variable included is the share of concessional debt to total debt (CONCDEB). Concessional debt is defined as lending extended by the IMF, Worldbank or other development banks at terms below the market terms (IMF, 2003). The concessionality is achieved, e.g., by interest rates below those available on the market and/or by grace periods. Foreign variables such as the US real interest rate (USINT) and the real growth rates of the OECD countries (GROWOECD) are also included to reflect the influence of the world economy. As in Milesi-Ferretti and Razin (1998), the current account, growth, investment and terms of trade variables are 3-years averages, in order to ensure consistency with the way reversals are measured.

3.3 Empirical Specifications

The baseline specification used for the analysis of current account reversals is a dynamic panel probit model of the form

$$y_{it}^* = x_{it}'\pi + \kappa y_{it-1} + e_{it}, \quad y_{it} = \mathbb{I}(y_{it}^* > 0), \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (3.2)$$

where $\mathbb{I}(y_{it}^* > 0)$ is an indicator function that transforms the latent continuous variable y_{it}^* for country i in year t into the binary variable y_{it} , indicating the occurrence of a current account reversal ($y_{it} = 1$), and x_{it} denotes the explanatory variables¹. The error term e_{it} is assumed to be normally distributed with zero mean and a fixed variance. Since equation (3.2) is only identified up to a positive multiplicative constant, a normalization condition will be required for each model variant. The vector x_{it} contains the observed macroeconomic, external, debt and foreign variables which might affect the incidence of a reversal. The lagged dependent variable on the right hand side is included to capture possible state dependence. It reflects the possibility that past current account crises could

¹Note that, different from the previous Chapter, x_i are not used to describe the nodes of an integral approximation.

lead to changes in institutional, political or economic factors affecting the probability of another reversal, and implies that the covariates in x_{it} have not only a contemporaneous but also a persistent effect on the probability of a reversal.

The most restrictive version of the panel probit assumes that the error e_{it} is independent across time and countries and imposes the restriction $\kappa = 0$. This produces the standard pooled probit estimator which ignores possible serial dependence and unobserved heterogeneity which cannot be attributed to the variables in x_{it} , but can be easily estimated and does not require numerical integration.

3.3.1 Random country-specific effects

Although easy to estimate, the pooled probit model does not account for country-specific differences such as property rights, tax systems, and other institutional factors which are difficult to control for, but affect a country's propensity to experience a current account reversal. In order to take these differences into account, fixed or random effect panel models could be used, but a fixed effect model based on country-specific dummy variables, such as the one used in the studies of Milesi-Ferretti and Razin (1998) and Edwards (2004a,b), requires the estimation of a large number of parameters, leading to a significant loss of degrees freedom. Furthermore, since the data set analyzed here includes countries that never experienced a reversal, for which the dependent variable does not vary, the ML-estimator does not exist (as shown in Table 3.7, the data set used includes countries that never experienced a reversal). This identification problem restricts the analysis to a random effect approach.

A prominent random effect model is that proposed by Butler and Moffitt (1982). It assumes the following specification for the error term in (3.2)

$$e_{it} = \tau_i + \epsilon_{it}, \quad \epsilon_{it} \sim \text{i.i.d.N}(0, 1), \quad \tau_i \sim \text{i.i.d.N}(0, \sigma_\tau^2). \quad (3.3)$$

The country-specific term τ_i captures possible permanent latent differences in the propensity to experience a reversal. Furthermore, it is assumed that τ_i and ϵ_{it} are independent from the variables included in x_{it} . If, however, x_{it} did contain variables reflecting countries' general susceptibility to current account crises, then τ_i would be correlated with x_{it} . Ignoring such correlation leads to inconsistent parameter estimates. Whence, the orthogonality condition between the random

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

effect and the included regressors should be tested. An independence test is discussed further below.

Notice that the time-invariant heterogeneity component τ_i implies a cross-period correlation of the error term e_{it} which is constant for all pairs of periods, and is given by $\text{corr}(e_{it}, e_{is}) = \sigma_\tau^2 / (\sigma_\tau^2 + 1)$ for $t \neq s$ (Greene, 2007). Additional potential sources of serial dependence are transitory country-specific differences in the propensity to experience a reversal leading to serial correlation in the error component ϵ_{it} of Equation (3.3). Furthermore, the intertemporal characteristics of the current account itself (Obstfeld and Rogoff, 1996), and the evidence of sluggish behavior of the trade balance (Baldwin and Krugman, 1989) and of foreign direct investments (Dixit, 1992) might introduce further serial dependence in ϵ_{it} . Whence, in addition to the Butler-Moffitt specification (3.2) and (3.3), it is assumed here that e_{it} includes a serially correlated idiosyncratic error component according to

$$e_{it} = \tau_i + \epsilon_{it}, \quad \epsilon_{it} = \rho\epsilon_{it-1} + \eta_{it}, \quad \eta_{it} \sim \text{i.i.d.N}(0, 1), \quad (3.4)$$

where τ_i and η_{it} are independent among each other and also from the variables included in x_{it} . In order to ensure stationarity $|\rho| < 1$ is assumed.

3.3.2 Random country- and time-specific effects

International capital markets, particularly those in emerging economies, appear volatile and subject to spillover effects. The currency crises of the 1990s and the way in which they rapidly spread across emerging markets including those rated as healthy economies by analysts and multilateral institutions, have brought interest in contagion effects (Edwards and Rigobon, 2002). A crisis in one country may lead investors to withdraw their investments from other markets without taking into account differences in economic fundamentals. In addition, a crisis in one economy can also affect the fundamentals of other countries through trade links and currency devaluations. Trading partners of a country in which a financial crisis has induced a sharp currency depreciation could experience a deterioration of the trade balance resulting from a decline in exports and an increase in imports (Corsetti et al., 1999). In the words of the former Managing Director of the IMF: “from the viewpoint of the international system, the devaluations in Asia will lead to large current account surpluses in those countries, damaging the

3.4 Maximum-Likelihood Estimation

competitive position of other countries and requiring them to run current account deficit” Fisher (1998).

Currency devaluations of countries that experience a crisis can often be seen as a *beggar-thy-neighbor* policy in the sense that they incite output growth and employment domestically at the expense of output growth, employment and current account deficit abroad (Corsetti et al., 1999). *Competitive devaluations* also happen in response to this process, as other economies may in turn try to avoid competitiveness loss through devaluations of their own currency. This appears to have happened during the East Asian crises in 1997 (Dornbusch et al., 2000).

The panel probit models introduced above do not account for such spillover effects since they ignore correlation across countries. In order to address this issue the following factor specification for the error e_{it} is also considered in the probit regression (3.2):

$$e_{it} = \tau_i + \xi_t + \epsilon_{it}, \quad \epsilon_{it} \sim \text{i.i.d.N}(0, 1), \quad (3.5)$$

with

$$\xi_t = \delta\xi_{t-1} + \nu_t, \quad \nu_t \sim \text{i.i.d.N}(0, \sigma_\xi^2), \quad (3.6)$$

where τ_i , ϵ_{it} and ν_t are mutually independent and independent from x_{it} . Furthermore, it is assumed that $|\delta| < 1$. The common dynamic factor ξ_t represents unobserved time-specific effects which induce correlation across countries, reflecting possible spillover effects. Note that this factor specification, which was also used in Liesenfeld and Richard (2008b) for a microeconomic application, resembles the linear panel models with a factor structure as discussed, e.g., in Baltagi (2005) and primarily used for the analysis of macroeconomic data.

3.4 Maximum-Likelihood Estimation

ML estimation of the simple pooled panel probit model is essentially the same as for a single equation probit model. The standard pooled probit does not take into account countries- or time-specific differences, which means it assumes that the $e_{i,t}$ s in equation (3.2) are independent across time and units and normally distributed with mean zero. The likelihood function of this model does not require integration with respect to any unobserved component, rendering

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

ML estimation of such models is straightforward. The loglikelihood of the pooled probit is given by:

$$\log L(\theta; \underline{y}, \underline{x}) = \sum_{i=1}^{NT} \{y_i \log [\Phi(x'_i \beta)] + (1 - y_i) \log [1 - \Phi(x'_i \beta)]\} \quad (3.7)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function (cdf) and NT is the total number of observations.

3.4.1 Random country-specific effects

Efficient parameter estimates can also be easily obtained for the probit model allowing for random country-specific effects given by (3.2) and (3.3). In particular, the event probabilities are represented by one dimensional integrals, which can be evaluated conveniently by means of a quadrature procedure. Let $\underline{y} = \{\{y_{it}\}_{t=1}^T\}_{i=1}^N$ and $\underline{x} = \{\{x_{it}\}_{t=1}^T\}_{i=1}^N$. Let θ denote the parameter vector to be estimated. The likelihood function for the Butler-Moffitt random effect model is then given by

$$L(\theta; \underline{y}, \underline{x}) = \prod_{i=1}^N \left\{ \int_{\mathbb{R}^1} \prod_{t=1}^T [\Phi_{it}^{y_{it}} (1 - \Phi_{it})^{(1-y_{it})}] \frac{1}{\sqrt{2\pi\sigma_\tau^2}} \exp \left\{ \frac{-\tau_i^2}{2\sigma_\tau^2} \right\} d\tau_i \right\}, \quad (3.8)$$

where $\Phi_{it} = \Phi(x'_{it}\pi + \kappa y_{it-1} + \tau_i)$, and again $\Phi(\cdot)$ represents the cdf of the standardized normal distribution. In the application below, the one dimensional integrals in τ_i are evaluated using the Gauss-Hermite quadrature rule (see Section 2.1.1.2).

Once the parameters have been estimated, the Gauss-Hermite procedure can also be used to obtain estimates of the random country-specific effects τ_i . Those estimates can serve as the basis for validating the imposed orthogonality conditions between the τ_i 's and the included regressors. In particular, the conditional expectation of τ_i given the sample information $(\underline{y}, \underline{x})$ is obtained as

$$\hat{\tau}_i = E(\tau_i | \underline{y}, \underline{x}; \theta) = \frac{\int_{\mathbb{R}} \tau_i h_i(\underline{y}_i, \tau_i | \underline{x}_i; \theta) d\tau_i}{\int_{\mathbb{R}} h_i(\underline{y}_i, \tau_i | \underline{x}_i; \theta) d\tau_i}, \quad (3.9)$$

where h_i denotes the joint conditional pdf of $\underline{y}_i = \{y_{it}\}_{t=1}^T$ and τ_i given $\underline{x}_i = \{x_{it}\}_{t=1}^T$, as defined by the integrand of the likelihood (3.8). For the evaluation of the numerator and denominator by Gauss-Hermite, the parameters θ are set to their ML-estimates. An auxiliary regression of the estimated random effects

3.4 Maximum-Likelihood Estimation

$\hat{\tau}_i$ against the time average of the explanatory variables provides a direct test of the validity of the orthogonality condition between τ_i and \underline{x}_i .

In contrast to the Butler-Moffitt model, the computation of the likelihood for the model (3.2) and (3.4) with country-specific effects and a serially correlated idiosyncratic error component requires the evaluation of $(T+1)$ -dimensional interdependent integrals, and that of the model (3.2), (3.5), and (3.6) with country-specific and time effects the evaluation of $(T+N)$ -dimensional integrals. As seen in Chapter 2, deterministic numerical integration procedures are not able to evaluate such high-dimensional integrals, and thus the estimation of these models have to rely on Monte Carlo approximations to the likelihood function.

3.4.2 AR(1) country-specific errors

The likelihood function of the panel probit model defined by equations (3.2) and (3.4) has the form $L(\theta; \underline{y}, \underline{x}) = \prod_{i=1}^N I_i(\theta)$, where I_i represents the likelihood contribution of country i . In the following, the likelihood function for a single country is derived, deleting the subscript i for the ease of notation. Let $\underline{\lambda}'_t = (\epsilon_t, \epsilon_{t-1}, \tau)$, $\underline{\lambda}'_0 = (\tau_0)$, $\underline{\lambda}'_1 = (\epsilon_1, \tau_i)$, and $\underline{\lambda}' = (\tau_i, \epsilon_1, \dots, \epsilon_T)$. Under the assumption that $\epsilon_0 = 0$, the likelihood contribution of a country is given by

$$I(\theta) = \int_{\mathbb{R}^{T+1}} \left[\prod_{t=1}^T \varphi_t(\underline{\lambda}_t) \right] f_\tau(\tau) d\underline{\lambda}, \quad (3.10)$$

with

$$\varphi_t(\underline{\lambda}_t) = \begin{cases} \mathbb{I}(\epsilon_t \in D_t) \phi(\epsilon_t - \rho \epsilon_{t-1}), & \text{if } t > 1 \\ \mathbb{I}(\epsilon_1 \in D_1) \phi(\epsilon_1), & \text{if } t = 1, \end{cases} \quad (3.11)$$

$$D_t = \begin{cases} [-(\mu_t + \tau), \infty), & \text{if } y_t = 1 \\ (-\infty, -(\mu_t + \tau)], & \text{if } y_t = 0, \end{cases} \quad (3.12)$$

where $\mu_t = x'_t \pi + \kappa y_{t-1}$, \mathbb{I} is the indicator function, and $\phi(\cdot)$ denotes the standardized Normal density. $I(\theta)$ will be evaluated by Importance Sampling under trajectories $\{\tilde{\underline{\lambda}}^{(j)}\}_{j=1}^S$ drawn from an importance sampler $m(\underline{\lambda}|\cdot)$. In order to simplify the sequential application of importance sampling, it proves convenient to rewrite (3.10) as

$$I(\theta) = \int_{\mathbb{R}^{T+1}} \prod_{t=0}^T \varphi_t(\underline{\lambda}_t) d\underline{\lambda}, \quad (3.13)$$

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

where $\varphi_0(\lambda_0) = f_\tau(\tau)$. Also $\underline{\lambda}'_t$ is partitioned into $(\epsilon_t, \underline{\eta}'_{t-1})$ with $\underline{\eta}'_{t-1} = (\epsilon_{t-1}, \tau_0)$ for $t > 1$, $\eta_0 = \lambda_0$ and $\eta_{-1} = \emptyset$.

3.4.2.1 Estimation via the GHK algorithm

The GHK simulator presented in Section 2.2.2.1 is an importance sampler developed to simulate multivariate normal rectangle probabilities (see Hajivassiliou et al., 1996), and as such it is well suited to simulate (3.13). It exploits the property of multivariate normal distribution that the conditional distribution of one element of $\underline{\epsilon}$ given a subvector of $\underline{\epsilon}$ is also normally distributed.

Note that in (3.11) the random vector $\underline{\epsilon}$ was already transformed to get a random vector with standard normal distribution using (3.4), and that the event probability can be rewritten as

$$I(\theta) = \text{Prob}(\epsilon_1 < \gamma_1 + \delta_1 \lambda_0, \epsilon_2 < \gamma_2 + \delta_2 \lambda_0 + \delta_2 \rho \epsilon_1, \dots, \epsilon_T < \gamma_T + \delta_T \lambda_0 + \delta_T \rho \epsilon_{T-1}),$$

which can be decomposed into a product of conditional event probabilities

$$I(\theta) = \prod_{t=1}^T \Phi(\gamma_t + \delta_t \lambda_0 + \delta_t \rho \epsilon_{t-1} | \underline{\eta}_{t-1}). \quad (3.14)$$

The GHK importance sampler is therefore

$$\begin{aligned} m(\underline{\epsilon}) &= \prod_{t=0}^T \frac{\varphi_t(\underline{\lambda}_t)}{\Phi(\gamma_t + \delta_t \lambda_0 + \delta_t \rho \epsilon_{t-1} | \underline{\eta}_{t-1})} \\ &= \prod_{t=0}^T m_t(\epsilon_t | \underline{\eta}_{t-1}), \end{aligned} \quad (3.15)$$

where $\Phi(\gamma_0 + \delta_0 \lambda_0 + \delta_0 \rho \epsilon_{-1} | \underline{\eta}_{-1}) \equiv 1$, and $m_0(\lambda_0) = f_\tau(\tau)$ for period $t = 0$. The GHK importance sampling representation of (3.13) is given by

$$\begin{aligned} I(\theta) &= \int_{\mathbb{R}^{T+1}} \prod_{t=0}^T \varphi_t(\underline{\lambda}_t) d\underline{\lambda} \\ &= \int_{\mathbb{R}^{T+1}} \prod_{t=0}^T \frac{\varphi_t(\underline{\lambda}_t)}{m_t(\epsilon_t | \underline{\eta}_{t-1})} \cdot m_t(\epsilon_t | \underline{\eta}_{t-1}) d\underline{\lambda} \\ &= \int_{\mathbb{R}^{T+1}} \prod_{t=1}^T \Phi(\gamma_t + \delta_t \lambda_0 + \delta_t \rho \epsilon_{t-1} | \underline{\eta}_{t-1}) \cdot m_t(\epsilon_t | \underline{\eta}_{t-1}) d\underline{\lambda}, \end{aligned} \quad (3.16)$$

3.4 Maximum-Likelihood Estimation

where $m_0(\lambda_0)$ and $\varphi_0(\lambda_0)$ canceled out in the last equality. The GHK-MC estimate of $I(\theta)$ is obtained as

$$\bar{I}_S(\theta) = \frac{1}{S} \sum_{j=1}^S \left[\prod_{t=1}^T \Phi(\gamma_t + \delta_t \tilde{\lambda}_0^{(j)} + \delta_t \rho \tilde{z}_{t-1}^{(j)}) \right], \quad (3.17)$$

where $\{\tilde{\lambda}^{(j)}\}_{j=1}^S$ denotes i.i.d. trajectories drawn from the sequential samplers $(m_0(\lambda_0), \{m_t(\epsilon_t | \eta_{t-1})\}_{t=1}^T)$.

The GHK procedure can also be used to obtain a MC estimate of the conditional expectation of $\tau | \underline{y}, \underline{x}$ according to (3.9). Under the panel probit model (3.2) and (3.4), the joint conditional pdf of \underline{y} and τ given \underline{x} takes the form

$$h(\underline{y}, \tau | \underline{x}; \theta) = \left[\int_{\mathbb{R}^T} \prod_{t=1}^T \varphi_t(\underline{\lambda}_t) d\epsilon_1 \cdots d\epsilon_T \right] f_\tau(\tau), \quad (3.18)$$

which is used in Equation (3.9) in order to produce MC-GHK estimates of the conditional expectation of τ .

3.4.2.2 Efficient Estimation via the EIS algorithm

EIS aims at constructing a sequence of auxiliary importance samplers to (3.13) of the form

$$m_t(\epsilon_t | \eta_{t-1}; a_t) = \frac{k_t(\underline{\lambda}_t; a_t)}{\chi_t(\eta_{t-1}; a_t)}, \quad t = 0, \dots, T, \quad (3.19)$$

with

$$\chi_t(\eta_{t-1}; a_t) = \int_{\mathbb{R}} k_t(\underline{\lambda}_t; a_t) d\epsilon_t, \quad (3.20)$$

where $\{k_t(\underline{\lambda}_t; a_t); a_t \in A_t\}$ denote a (pre-selected) class of auxiliary parametric density kernels with $\chi_t(\eta_{t-1}; a_t)$ as analytical integrating factor in ϵ_t given (η_{t-1}, a_t) . The integral in (3.13) is rewritten as

$$I(\theta) = \chi_0(a_0) \int_{\mathbb{R}^{T+1}} \prod_{t=0}^T \left[\frac{\varphi_t(\underline{\lambda}_t) \chi_{t+1}(\eta_t; a_{t+1})}{k_t(\underline{\lambda}_t; a_t)} \right] m_t(\epsilon_t | \eta_{t-1}; a_t) d\underline{\lambda} \quad (3.21)$$

with $\chi_{T+1}(\cdot) \equiv 1$. The backward transfer of the integrating factors $\chi_{t+1}(\cdot)$ constitutes the cornerstone of sequential EIS and is meant to capture as closely as possible the dynamics of the underlying process. As it was pointed out in Section 2.2.2.1, it is precisely the lack of such transfers which explains the inefficiency

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

of the GHK procedure in (large dimensional) interdependent truncated integrals. Under (3.21), an EIS-MC estimate of $I(\theta)$ is given by

$$\bar{I}_S(\theta) = \chi_0(a_0) \frac{1}{S} \sum_{j=1}^S \left[\prod_{t=0}^T \frac{\varphi_t(\tilde{\lambda}_t^{(j)}) \chi_{t+1}(\tilde{\eta}_t^{(j)}; a_{t+1})}{k_t(\tilde{\lambda}_t^{(j)}; a_t)} \right], \quad (3.22)$$

where $\{\tilde{\lambda}^{(j)} = \{\tilde{\lambda}_t^{(j)}\}_{t=0}^T\}_{j=1}^S$ denote S i.i.d. trajectories drawn from the auxiliary sampler

$$m(\underline{\lambda}|a) = \prod_{t=0}^T m_t(\epsilon_t | \underline{\eta}_{t-1}; a_t), \quad a = (a_0, \dots, a_T) \in A = \times_{t=0}^T A_t. \quad (3.23)$$

That is to say, $\tilde{\epsilon}_t^{(j)}$ is drawn from $m_t(\epsilon_t | \tilde{\eta}_{t-1}^{(j)}; a_t)$ for $t = 1, \dots, T$, and $\tilde{\lambda}_0^j$ is drawn from $m_0(\lambda_0, a_0)$. An efficient importance sampler is one which minimizes the MC sampling variances of the ratios $\varphi_t \chi_{t+1} / k_t$ under such draws. Since $m_t(\cdot; a_t)$ depends itself upon a_t , efficient a_t values are obtained as solutions of the following fixed point sequences of back-recursive auxiliary least squares (LS) problems:

$$(\hat{c}_t^{(k+1)}, \hat{a}_t^{(k+1)}) = \arg \min_{c_t, a_t} \sum_{j=1}^S \left\{ \ln \left[\varphi_t(\tilde{\lambda}_t^{(k,j)}) \cdot \chi_{t+1}(\tilde{\eta}_t^{(k,j)}; \hat{a}_{t+1}^{(k)}) \right] - c_t - \ln k_t(\tilde{\lambda}_t^{(k,j)}; a_t) \right\}^2, \quad (3.24)$$

for $t = T, T-1, \dots, 0$, where $\{\tilde{\lambda}^{(k,j)} = \{\tilde{\lambda}_t^{(k,j)}\}_{t=0}^T\}_{j=1}^S$ denote trajectories drawn from $m(\underline{\lambda} | \hat{a}^{(k)})$. Convergence to a fixed point solution typically requires 3 to 5 iterations for reasonably well-behaved applications. See Richard and Zhang (2007) or Section 2.2.3 for details. To initialize the search for \hat{a}_t , the starting values a^0 implied by the GHK importance sampler given in (3.15) are used.

Two additional key components of this EIS algorithm are as follows: (i) The kernel $k_t(\underline{\lambda}_t; a_t)$ has to approximate the ratio $\varphi_t(\underline{\lambda}_t) \cdot \chi_{t+1}(\underline{\eta}_t; a_{t+1})$ with respect to $\underline{\lambda}_t$, not just ϵ_t in order to capture the interdependence across the ϵ_t 's. (There is no revisiting of period t once \hat{a}_t has been found); (ii) All trajectories $\{\tilde{\lambda}^{(k,j)}\}_{j=1}^S$ have to be obtained by transformation of a single set of Common Random Numbers (CRNs) $\{\tilde{u}^{(j)}\}_{j=1}^S$ pre-drawn from a canonical distribution, i.e. one which does not depend on a . In the present case the CRNs consist of draws from a uniform distribution to be transformed into (truncated) Gaussian draws from $m_t(\epsilon_t | \tilde{\eta}_{t-1}; \hat{a}_t)$ (see Appendix 1 of this Chapter).

3.4 Maximum-Likelihood Estimation

Next, the specific application of EIS to the likelihood integral defined by Equation (3.10) is discussed. In the present section, only the heuristic of such implementation is presented. Full details are given in Appendix 1, where it is shown that the EIS problem in (3.24) actually reduces to a univariate EIS (instead of a trivariate one in $\underline{\lambda}_t$) by taking full advantage of the particular structure of $\chi_{t+1}(\cdot)$.

Note first that the period- t integrand in Equation (3.10) includes a (truncated) Gaussian kernel. Therefore, it appears appropriate to select a Gaussian kernel for $k_t(\underline{\lambda}_t; a_t)$, a choice further supported by the fact that it will be demonstrated that $\chi_t(\underline{\eta}_t; a_t)$ then takes the form of a Gaussian kernel times a probability. Moreover, the selection of a Gaussian kernel enables the use of the fact that the class of Gaussian kernels in $\underline{\lambda}_t$ is closed under multiplication (see DeGroot, 2004, or Section 2.2.3.1). Therefore, k_t is specified as the following product

$$k_t(\underline{\lambda}_t; a_t) = \varphi_t(\underline{\lambda}_t) \cdot k_{0,t}(\underline{\lambda}_t; a_t), \quad (3.25)$$

where $k_{0,t}$ is itself a Gaussian kernel in $\underline{\lambda}_t$. It immediately follows that $\varphi_t \cdot \chi_{t+1}/k_t \equiv \chi_{t+1}/k_{0,t}$ so that φ_t cancels out in the auxiliary EIS-LS optimization problem as defined in Equation (3.24). Under specification (3.25), the standard EIS implementation follows, noting that $\lambda_0 = \tau$ is present in all $T + 1$ factors of the integrand. Whence, one should proceed as follows:

(i) Regroup all terms in k_t which only depend on λ_0 . Let the corresponding factorization be denoted as

$$k_t(\underline{\lambda}_t; a_t) = k_{1,t}(\underline{\lambda}_t; a_t) \cdot k_{3,t}(\lambda_0; a_t); \quad (3.26)$$

(ii) Let $\chi_{1,t}$ denote the integral of $k_{1,t}$ w.r.t. ϵ_t such that

$$\chi_{1,t}(\underline{\eta}_{t-1}; a_t) = \int_{\mathbb{R}} k_{1,t}(\underline{\lambda}_t; a_t) d\epsilon_t. \quad (3.27)$$

Note that this integral is truncated to D_t^* due to the indicator function $\mathbb{I}(\underline{\epsilon} \in D)$ which is included in φ and, therefore, in $k_{1,t}$. It follows that $\chi_{1,t}$ takes the form of a Gaussian kernel in $\underline{\eta}_{t-1}$ times the probability that $\epsilon_t \in D_t^*$ conditional on $\underline{\eta}_{t-1}$, say

$$\chi_{1,t}(\underline{\eta}_{t-1}; a_t) = \Phi(\alpha_t + \beta_t' \underline{\eta}_{t-1}) \cdot k_{2,t}(\underline{\eta}_{t-1}; a_t), \quad (3.28)$$

where (α_t, β_t') are appropriate functions of a_t and the data.

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

It follows from Equations (3.26) to (3.28) that the integral of k_t w.r.t. ϵ_t is of the form

$$\chi_t(\underline{\eta}_{t-1}; a_t) = k_{3,t}(\lambda_0; a_t) [\Phi(\alpha_t + \beta'_t \underline{\eta}_{t-1}) \cdot k_{2,t}(\underline{\eta}_{t-1}; a_t)]. \quad (3.29)$$

In direct application of the backward transfer of integrating factors associated with sequential EIS, the factor $k_{3,t}$ is transferred back directly into the period $t = 0$ integral while the two factors between brackets are transferred back into the period $t - 1$ integral. Full details are provided in Appendix 1 of this Chapter.

The EIS procedure can also be used to obtain an accurate MC-estimate of the conditional expectation of $\tau | \underline{y}, \underline{x}$ according to Equation (3.9). Under the panel probit model (3.2) and (3.4), the joint conditional pdf of \underline{y} and τ given \underline{x} takes the form

$$h(\underline{y}, \tau | \underline{x}; \theta) = \left[\int_{\mathbb{R}^T} \prod_{t=1}^T \varphi_t(\underline{\lambda}_t) d\epsilon_1 \cdots d\epsilon_T \right] f_\tau(\tau), \quad (3.30)$$

which is used in Equation (3.9) in order to produce MC-EIS estimates of the conditional expectation of τ .

This heuristic presentation of the EIS application to the panel probit model defined by Equations (3.2) and (3.4) is concluded with two important comments. Firstly, as mentioned above, the MC procedure most frequently used to compute choice probabilities is the GHK technique already introduced above. It is also an importance sampling procedure but it selects φ_t itself as the auxiliary period- t kernel.

Note that the GHK importance sampler actually belongs to the class of auxiliary EIS samplers introduced in Equation (3.25) since it amounts to selecting a diffuse $k_{0,t} \propto 1$. Therefore, it is inefficient within this class. Results of a MC experiment provided in Appendix 2 highlight the inefficiency of GHK in the context of the particular model analyzed here. A broader MC investigation of the relative inefficiency of GHK relative to EIS can be found in Liesenfeld and Richard (2007).

Secondly, Zhang and Lee (2004) offer a MC comparison between GHK and AIS (Accelerated Importance Sampling), where AIS denotes an earlier version of EIS, in the context of the same model as that defined in Equations (3.2) and (3.4). In apparent contrast with the results presented below, they find that GHK performs essentially as well as AIS, except for longer panels. However, their AIS algorithm

3.4 Maximum-Likelihood Estimation

differs from the EIS implementation presented above in several critical aspects. Most importantly, it ignores the factorization (3.26) and has no direct transfer of integrating factors depending solely on $\lambda_0 = \tau$. This turns out to be a major source of inefficiency since critical information relative to τ is then filtered through the full sequence of AIS approximations instead of being transferred directly back into the $t = 0$ integral. Moreover, it would appear that AIS optimizations are not iterated toward fixed point solutions as in Equation (3.24), nor does AIS relies upon CRNs. Last but not least, the simulation results in Zhang and Lee are based upon i.i.d. replications of the actual sampling process (with no indications as to how the auxiliary AIS draws are produced across replications). Whence, the standard deviations and/or mean squared errors they report are measures of conventional statistical dispersions with no indications of numerical accuracy. In contrast, the results reported below are based upon i.i.d. replications of the CRNs for a given sample and are, therefore, meant to measure only numerical accuracy. See Richard and Zhang (2007) for a discussion of the importance of these additional EIS refinements.

3.4.3 AR(1) time-specific effects

The likelihood function for the random effect panel model consisting of Equations (3.2), (3.5), and (3.6) is given by

$$L(\theta; \underline{y}, \underline{x}) = \int_{\mathbb{R}^{T+N}} \prod_{i=1}^N \prod_{t=1}^T [\Phi(z_{it})]^{y_{it}} [1 - \Phi(z_{it})]^{(1-y_{it})} p(\underline{\tau}, \underline{\xi}) d\underline{\tau}, d\underline{\xi}, \quad (3.31)$$

where $\underline{\xi} = \{\xi_t\}_{t=1}^T$, $\underline{\tau} = \{\tau_i\}_{i=1}^N$, $z_{it} = x'_{it}\pi + \kappa y_{it-1} + \tau_i + \xi_t$, and $p(\underline{\tau}, \underline{\xi})$ denotes the joint density of the unobserved effects. The presence of a time effect ξ_t common to all countries prevents the factorization of the likelihood function into a product of integrals for each individual country. The joint density of $(\underline{\tau}, \underline{\xi})$ is assumed to be proportional to

$$p(\underline{\tau}, \underline{\xi}) \propto \sigma_\tau^{-N} \exp \left\{ -\frac{1}{2\sigma_\tau^2} \tau_i^2 \right\} |H_\xi|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \underline{\xi}' H_\xi \underline{\xi} \right\}, \quad (3.32)$$

where H_ξ denotes the precision matrix of $\underline{\xi}$. See Richard (1977) for analytical expressions of H_ξ under alternative initial conditions, including stationarity. Conditionally on $\underline{\xi}$, one could apply GHK to each country individually, though

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

Gauss-Hermite would likely be more efficient for these univariate integrals in τ_i . One would then be left with a complicated T -dimensional integral in $\underline{\xi}$. In contrast, EIS can be applied to the likelihood function (3.31) in a way which effectively captures the complex *a posteriori* interdependence between $\underline{\tau}$ and $\underline{\xi}$.

The integrand in Equation (3.31) is first factorized as follows

$$L(\theta; \underline{y}, \underline{x}) = \int_{\mathbb{R}^{T+N}} \phi_0(\underline{\xi}) \prod_{i=1}^N \phi_i(\tau_i, \underline{\xi}) d\underline{\tau} d\underline{\xi}, \quad (3.33)$$

where

$$\phi_0(\underline{\xi}) \propto |H_\xi|^{\frac{1}{2}} \exp \left[-\frac{1}{2} \xi' H_\xi \xi \right] \quad (3.34)$$

$$\phi_i(\tau_i, \underline{\xi}) \propto \sigma_\tau^{-1} \exp \left[-\frac{1}{2\sigma_\tau^2} \tau_i^2 \right] \prod_{t=1}^T [\Phi(z_{it})]^{y_{it}} [1 - \Phi(z_{it})]^{(1-y_{it})}. \quad (3.35)$$

It is critical that the EIS sampler $m(\underline{\tau}, \underline{\xi}; a)$ fully reflects the interdependence structure of the posterior density of $(\underline{\tau}, \underline{\xi})$ which is proportional to the integrand in Equation (3.33). Specifically, the τ_i 's are independent from one another *conditionally* on $\underline{\xi}$ but are individually linked to the full $\underline{\xi}$ -vector. Accordingly, the auxiliary sampler is factorized as

$$m(\underline{\tau}, \underline{\xi}; a) = m_0(\underline{\xi}; a_0) \prod_{i=1}^N m_i(\tau_i | \underline{\xi}; a_i). \quad (3.36)$$

The corresponding kernels $\{k_i(\tau_i, \underline{\xi}; a_i)\}_{i=1}^N$ and $k_0(\underline{\xi}; a_0)$ are specified as *joint* Gaussian kernels in $(\tau_i, \underline{\xi})$ and $\underline{\xi}$, respectively. Significant simplifications follow from the particular form of the integrand in Equation (3.33). First, note that $\ln \phi_i$ is given by

$$\ln \phi_i(\tau_i, \underline{\xi}) \propto -\frac{1}{2} \frac{\tau_i^2}{\sigma_\tau^2} + \sum_{t=1}^T \ln \{ [\Phi(z_{it})]^{y_{it}} [1 - \Phi(z_{it})]^{(1-y_{it})} \}. \quad (3.37)$$

Each factor in the sum depends only on a single z_{it} . Therefore, $\ln k_i$ is specified as follows

$$\ln k_i(\tau_i, \underline{\xi}; a_i) = -\frac{1}{2} \left[\frac{\tau_i^2}{\sigma_\tau^2} + \sum_{t=1}^T (\alpha_{i,t} z_{i,t}^2 + 2\beta_{i,t} z_{i,t}) \right] \quad (3.38)$$

for a total of $2 \cdot T$ auxiliary parameters plus the intercept. It follows that, at the cost of standard algebraic operations $\chi_i(\underline{\xi}; a_i)$ (i.e. the integrating constant for

k_i) is itself a Gaussian kernel in $\underline{\xi}$. Whence, the product $\phi_0(\underline{\xi}) \cdot \prod_{i=1}^N \chi_i(\underline{\xi}; a_i)$ is a Gaussian kernel and requires no further adjustment (an interesting example of perfect fit in an EIS auxiliary regression).

Estimates for the unobserved random effects are obtained, in the same way as under the panel model without time effects, as by-products of the EIS computation of the likelihood. In particular, the conditional expectation of the vector of random effects $\underline{v} = (\underline{\tau}, \underline{\xi})$ given the sample information has the form

$$\hat{\underline{v}} = \text{E}(\underline{v} | \underline{y}, \underline{x}; \theta) = \frac{\int_{\mathbb{R}^{N+T}} \underline{v} h(\underline{y}, \underline{v} | \underline{x}; \theta) d\underline{v}}{\int_{\mathbb{R}^{N+T}} h(\underline{y}, \underline{v} | \underline{x}; \theta) d\underline{v}}, \quad (3.39)$$

where h denotes the joint conditional pdf of \underline{y} and \underline{v} given \underline{x} which is given by the integrand of the likelihood function (3.31).

3.5 Empirical Results

3.5.1 A note on normalization

In Equations (3.3) to (3.6), the standard practice of normalizing the probit equation (3.2) by setting the variance of the residual innovations equal to 1 is followed. Therefore the variances of the composite error term e_{it} differ across models, implying corresponding differences in the implicit normalization rule. The variances of e_{it} under the different specifications are given by

$$\begin{aligned} \text{Equation (3.3)} & : & \sigma_e^2 &= 1 + \sigma_\tau^2 \\ \text{Equation (3.4)} & : & \sigma_e^2 &= \frac{1}{1 - \rho^2} + \sigma_\tau^2 \\ \text{Equations (3.5)+(3.6)} & : & \sigma_e^2 &= 1 + \sigma_\tau^2 + \frac{\sigma_\xi^2}{1 - \delta^2}. \end{aligned}$$

These differences affect comparisons between estimated coefficients across models (but not between estimated probabilities). The implied correction factors do not exceed 11% for the results reported below. For the ease of comparison the estimated standard deviation σ_e of e_{it} for each model will be reported.

3.5.2 Model 1: Pooled Probit

The ML estimate of the pooled probit model (3.2) under the assumption that the errors are independent across time and countries together with the corresponding

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

estimated partial effects of the explanatory variables on the probability of a current account reversal are presented in Table 3.1. The results for the static model ($\kappa = 0$) are reported in the left columns and those of the dynamic specification including the lagged dependent in the right columns.

The parameter estimates of the covariates x_{it} are all in line with the results in the empirical literature on current account crises (see Edwards, 2004a,b; Milesi-Ferretti and Razin, 1998). Sharp reductions of the current-account deficit are more likely in countries with a high current account deficits (AVGCA) and with higher government expenditures (GOV). The significant effect of the current account deficit level is consistent with a need for sharp corrections in the trade balance to ensure that the country remains solvent. Interpreting current account as a constraint on expenditures, the positive impact of government expenditure on the reversal probability can be attributed to fact that an increase of government expenditures leads to a deterioration of the current account. However, the inclusion of the lagged dependent variable reduces this effect and makes it non significant. This suggests that government expenditures might capture some omitted serial dependence under the static specification. The marginal effect of foreign reserve (RES) is negative and significant which suggests that low levels of reserves make it more difficult to sustain a large trade imbalance and may also reduce foreign investors' willingness to lend (Milesi-Ferretti and Razin, 1998). Also, reversals seem to be less common in countries with a high share of concessional debt (CONCDEB). A possible explanation for this is that concessional debt flows are typically less volatile and less likely to be sharply reversed than flows of other debt and equity instruments used to finance current account deficits like foreign portfolio investments (Milesi-Ferretti and Razin, 1998). Furthermore, concessional debt flows are typically higher in poor countries that have difficulties reducing external imbalances and serving external obligations leading to a lower probability of observing a reversal for those countries. In fact, low-income countries tend to run large current account deficits financed by official grants and loans. Finally, countries with a lower degree of openness (OPEN), weaker terms of trade (AVGTT) and higher GDP growth (AVGGROW) seem to face higher probabilities of reversals, especially when growth rate in OECD countries (GROWOECD) and/or US interest rate (USINT) are higher - though none of these these coefficients are statistically significant.

3.5 Empirical Results

Note that the size of the estimated marginal effects for the significant economic covariates on the probability of reversals are typically fairly small, ranging from 0.004 to 0.026. Nevertheless, they are far from being negligible when applied to the low unconditional probability of experiencing a reversal which is approximately 0.1.

The inclusion of the lagged current account reversal variable substantially improves the fit of the model as indicated by the significant increase of the maximized log-likelihood value. The estimated coefficient κ measuring the impact of the lagged dependent state variable is positive and significant at the 1% significance level with a large estimated partial effect of 0.21. This suggests that a current account reversal significantly increases the probability of a further reversal the following year. This would be consistent with the hypothesis that reversal processes stretch over more than a year due to slow adjustments in international trade flows (see Himarios, 1989; Junz and Rhomberg, 1973).

In order to analyze the dynamic effects of a covariate x_{itk} implied by the model with lagged dependent variable we use the sample average of the l -step ahead marginal effect, i.e.,

$$\frac{1}{N(T-l)} \sum_{i=1}^T \sum_{t=1}^{T-l} \partial_{x_{itk}} p(y_{it+l} = 1 | x_{it+l}, \dots, y_{it-1}), \quad l = 1, 2, \dots \quad (3.40)$$

The probability $p(y_{it+l} = 1 | x_{it+l}, \dots, y_{it-1})$ is obtained by considering the event tree associated with all possible y_{it} -trajectories starting in period t and ending in period $t+l$ with $y_{it+l} = 1$. Analogously, the dynamic effect of the state variable is measured by

$$\frac{1}{N(T-l)} \sum_{i=1}^T \sum_{t=1}^{T-l} [p(y_{it+l} = 1 | x_{it+l}, \dots, y_{it-1} = 1) - p(y_{it+l} = 1 | x_{it+l}, \dots, y_{it-1} = 0)], \quad l = 1, 2, \dots \quad (3.41)$$

The upper left panel of Figure 3.1 plots the dynamic marginal effects for the significant covariates (AVGCA, RES, CONCDEB) and the lagged state variable for $l = 1, \dots, 4$, respectively. It reveals substantial long-run effects of the state variable, whereby the occurrence of a current account reversal increases a country's propensity to experience further large reductions in the current account in subsequent years. This effect appears to stretch over a two-to-three-year period.

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

Table 3.1: ML estimates of Model 1: Pooled probit

Variable	Static Model		Dynamic Model	
	Estimate	Marg. Eff.	Estimate	Marg. Eff.
Constant	-1.993*** (0.474)		-1.955*** (0.481)	
AVGCA	-0.060*** (0.012)	-0.009	-0.060*** (0.012)	-0.009
AVGGROW	0.008 (0.021)	0.001	0.010 (0.021)	0.001
AVGINV	-0.002 (0.010)	-0.0003	0.00002 (0.011)	0.000003
AVGTT	-0.108 (0.066)	-0.017	-0.103 (0.064)	-0.015
GOV	0.026** (0.012)	0.004	0.019 (0.011)	0.003
OT	-0.011 (0.010)	-0.002	-0.011 (0.010)	-0.002
OPEN	-0.058 (0.087)	-0.009	-0.064 (0.086)	-0.009
USINT	0.108 (0.073)	0.017	0.081 (0.073)	0.012
GROWOECD	0.084 (0.086)	0.013	0.075 (0.084)	0.011
INTPAY	0.024 (0.029)	0.004	0.013 (0.030)	0.002
RES	-0.074** (0.030)	-0.011	-0.067** (0.029)	-0.010
CONCDEB	-0.165** (0.068)	-0.026	-0.149** (0.067)	-0.021
κ			0.998*** (0.147)	0.213
Log-likelihood	-276.13		-254.57	

Note: The estimated model is given by Equation (3.2) assuming that the errors are independent across countries and time. The asymptotic standard errors are given in parentheses and obtained from the inverse Hessian. *, **, and *** indicates statistical significance at the 10%, 5% and 1% significance level.

This would be in line with the result of Himarios (1989) showing that changes in trade flows triggered by currency devaluations often used to correct the trade balance are distributed over a time span of about two or three years. However, note that this long-run state dependence does not translate into significant long-run effects of the covariates AVGCA, RES, and CONCDEB which is consistent with the fact that their contemporaneous effects reported in Table 3.1 are already fairly small.

3.5.3 Model 2: Random country-specific effects

Table 3.2 reports the estimates of the Butler-Moffitt model (3.2) and (3.3), which includes random country specific effects, leading to equicorrelated errors across time periods. The ML-estimates are obtained using a 20-points Gauss Hermite quadrature. The estimate of the coefficient σ_τ indicates that only 3% of the total variation in the latent error is due to unobserved country-specific heterogeneity and this effect is not statistically significant. Furthermore, the maximized log-likelihood of the random effect model is only marginally larger than that of the dynamic pooled probit model with a likelihood-ratio (LR) test statistic of 0.2. Hence, there is no evidence in favor of the random effect specification for time-invariant differences of institutional, political, and economic factors across countries. Actually, the marginal effects as well as the predicted dynamic effects (see, upper right panel of Figure 3.1) obtained under the random country-specific effect model are very similar to those for the dynamic pooled model.

The estimated probit model with random effects assumes that τ_i is independent of x_{it} . If this were not correct, the parameter estimates would be inconsistent. In order to check this assumption the following auxiliary regression was used:

$$\hat{\tau}_i = \psi_0 + \bar{x}_i' \psi_1 + \zeta_i, \quad i = 1, \dots, n, \quad (3.42)$$

where the vector \bar{x}_i contains the mean values of the x_{it} -variables (except for the US interest rate and the OECD growth rate) over time. The value of the F -statistic for the null $\psi_1 = 0$ is 1.85 with critical values of 2.03 and 1.73 for the 5% and 10% significance levels. Whence, evidence that τ_i might be correlated with \bar{x}_i is inconclusive.

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

Table 3.2: ML estimates of Model 2: Random country-specific effects

Variable	Estimate	Marg. Eff.
Constant	-1.880*** (0.534)	
AVGCA	-0.064*** (0.015)	-0.009
AVGGROW	0.010 (0.021)	0.001
AVGINV	-0.0001 (0.011)	-0.00001
AVGTT	-0.122 (0.084)	-0.017
GOV	0.018 (0.012)	0.003
OT	-0.011 (0.011)	-0.002
OPEN	-0.069 (0.093)	-0.010
USINT	0.083 (0.075)	0.012
GROWOECD	0.073 (0.090)	0.010
INTPAY	0.014 (0.031)	0.002
RES	-0.073** (0.035)	-0.010
CONCDEB	-0.159** (0.078)	-0.023
κ	0.982*** (0.154)	0.206
σ_τ	0.162 (0.210)	
σ_e	1.013	
Log-likelihood	-254.47	
LR-statistic for $H_0 : \sigma_\tau = 0$	0.20	
F -statistic for exogeneity	1.85	

Note: The estimated model is given by (3.2) and (3.3). The asymptotic standard errors are in parentheses. *, **, and *** indicates statistical significance at the 10%, 5% and 1% significance level. The 1% and 5% percent critical values of the LR-statistic for $H_0 : \sigma_\tau = 0$ are 5.41 and 2.71. The 1% and 5% percent critical values of the F -statistic (t-statistic) are 2.71 and 2.03 (2.68 and 2.01).

3.5.4 Model 3: AR(1) country-specific errors

The dynamic random effect model with serially correlated idiosyncratic errors as specified by Equations (3.2) and (3.4) allows for a third source of serial dependence in addition to state dependence and time-invariant unobserved heterogeneity. It ought to capture possible serially correlated shocks associated with regional political changes or conflicts and persistent local macroeconomic events like commodity price shocks. The ML-EIS estimation results based on $S = 100$ EIS draws are given in the left columns of Table 3.3. The MC (numerical) standard deviations are computed from 20 ML-EIS estimations each of them based on a different i.i.d. set of CRNs. They are much smaller than the corresponding asymptotic (statistical) standard deviations indicating that the ML-EIS results are numerically very accurate.

The estimation results indicate that the inclusion of a transitory idiosyncratic error component has significant effects on the dynamic structure of the model but only a slight impact on the marginal effects of the x_{it} -variables, which remain typically very close to those of the pure random country-specific effect model in Table 3.2. An exception is the effect of the terms of trade (AVGTT) which becomes significant at the 10% level. Also, while the parameter σ_τ governing the time-invariant heterogeneity remains statistically insignificant, the estimated coefficient κ associated with the lagged dependent variable and its partial effect are now much smaller. This leads to a substantial attenuation of the long-run effect of the lagged state variable (see lower left panel of Figure 3.1). The estimate of the persistence parameter of the AR(1) error component ρ equals 0.35 and is statistically significant at the 10% level. However, the corresponding LR-statistic equals 2.40 and is not significant. Hence, despite its impact on the dynamic structure of the model, the inclusion of an AR(1) error component does not significantly improve the overall fit.

Since a lagged dependent variable and a country-specific AR(1) error component can generate similar looking patterns of persistence in the dependent variable, these results suggest that the AR(1) error captures some of the serial dependence which is captured by the lagged dependent variable under the pooled probit and the pure random country-specific effect model. However, the small likelihood improvement obtained by the inclusion of an AR(1) error together with the fairly large standard errors of the estimates for κ and ρ suggest that

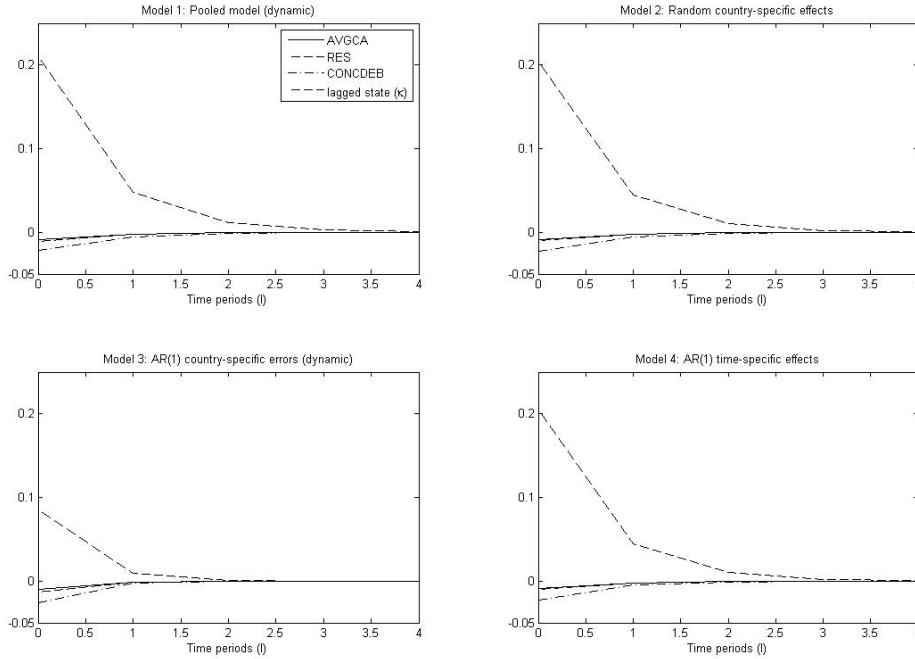
3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

Table 3.3: ML-EIS estimates of Model 3: AR(1) country-specific errors.

Variable	Dynamic		Static	
	Estimate	Marg. Eff.	Estimate	Marg. Eff.
Constant	-1.795*** (0.567)		-1.512** (0.677)	
AVGCA	-0.072*** (0.018)	-0.010	-0.087*** (0.021)	-0.012
AVGGROW	0.007 (0.024)	0.001	0.0001 (0.027)	0.00001
AVGINV	0.004 (0.013)	0.001	0.010 (0.017)	0.001
AVGTT	-0.161* (0.093)	-0.022	-0.251** (0.116)	-0.034
GOV	0.018 (0.014)	0.002	0.016 (0.018)	0.002
OT	-0.010 (0.012)	-0.001	-0.009 (0.014)	-0.001
OPEN	-0.108 (0.109)	-0.015	-0.175 (0.136)	-0.023
USINT	0.097 (0.075)	0.013	0.119 (0.082)	0.016
GROWOECD	0.057 (0.087)	0.008	0.038 (0.095)	0.005
INTPAY	0.029 (0.035)	0.004	0.045 (0.037)	0.006
RES	-0.097** (0.046)	-0.013	-0.143*** (0.054)	-0.019
CONCDEB	-0.190** (0.088)	-0.026	-0.261*** (0.099)	-0.035
κ	0.520* (0.297)	0.088		
σ_τ	0.142 (0.322)		0.194 (0.403)	
ρ	0.349* (0.198)		0.590*** (0.090)	
σ_e	1.077		1.254	
Log-likelihood	-253.27		-255.17	
LR-statistic for $H_0 : \rho = 0$	2.40		36.65	
F -statistic for exogeneity of x_{it}	2.16		2.54	
t -statistic for exogeneity of y_{i0}	-1.84			

Note: The estimated model is given by Equations (3.2) and (3.4). The ML-EIS estimation are based on a MC sample size of $S = 100$. The asymptotic standard errors are given in parentheses and obtained from the inverse Hessian. *, **, and *** indicates statistical significance at the 10%, 5% and 1% significance level. The 1%, 5%, and 10% percent critical values of the LR-statistic for $H_0 : \rho = 0$ are 6.63, 3.84, and 2.71. The 1% and 5% critical values of the F -statistic (t -statistic) are 2.71 and 2.03 (2.68 and 2.01).

Figure 3.1: Average ℓ -Step Ahead Marginal Effects



Average ℓ -step ahead marginal effects of the covariates AVGCA, RES, CONCDEB and the lagged binary state variable computed according to Equations (3.40) and (3.41).

the model has difficulties separating these two sources of serial dependence. In order to verify this conjecture, the model with the AR(1) country-specific error component without state-dependence was estimated. The ML-EIS results are provided in the right columns of Table 4 and confirm our conjecture. In fact, the estimated AR coefficient ρ increases to 0.59 and is now highly significant according to both the t- and LR-test statistics, while the maximized likelihood value are not significantly different from those obtained for the models including either state-dependence only (Table 3.2) or both state-dependence and an AR error component (left columns of Table 3.3).

All in all, our results indicate that the data are ambiguous on the question of whether the observed persistence in current account reversals is due to state

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

dependence associated with the hypothesis of slow adjustments in international trade flows or due to serially correlated country-specific shocks related to local political or macroeconomic events.

For the purpose of comparison, the random effect model with serially correlated errors is re-estimated using the standard GHK simulator based on the same simulation sample size as used for EIS ($S = 100$). The results, which are summarized in the right columns of Table 3.4, reveal that the parameter estimates obtained using GHK exhibit significantly larger MC standard errors than those obtained under EIS. Moreover, while the parameter estimates for the explanatory variables are generally similar for both procedures, the estimates of the parameters governing the dynamics of current account reversals (κ , σ_τ , ρ) are noticeably different. In particular, the ML-GHK estimates of σ_τ and ρ are smaller than their ML-EIS counterparts. This is fully in line with the results of the MC study of Lee (1997) indicating that the ML-GHK estimator exhibits a downward bias for the persistence parameter of the idiosyncratic error as well as for the variation parameter of the unobserved heterogeneity.

3.5.5 Model 4: AR(1) time-specific effects

Table 3.5 presents the ML estimation results of the panel model (3.2), (3.5), and (3.6), allowing for unobserved random effects in both dimensions designed to capture spill-over effects and/or global shocks common to all countries. The ML-EIS estimation was performed with a simulation sample size of $S = 100$. The MC standard errors reported illustrate how efficiently EIS approximates the $T + N$ integral in Equation (3.31).

The estimated marginal effects for all explanatory x_{it} -variables and the estimated variance parameter σ_τ of the time-invariant heterogeneity are very similar to those obtained under the models discussed above. Here again, there is no conclusive evidence for correlation between τ_i and (\bar{x}_i, y_{i0}) . The results show a large and highly significant state-dependence effect similar to that found under the pure random country-specific effect model in Table 3.2. The variance parameter of the time factor σ_ξ and its autoregressive parameter δ are both highly significant, indicating that there are significant common dynamic time-specific effects in addition to state dependence. Hence, in contrast to the specification with state dependence and an AR country-specific error component, the model

3.5 Empirical Results

Table 3.4: Model 3: Comparison between GHK and EIS ML estimation

Variable	ML-EIS			ML-GHK		
	Est.	Asy. s.e.	MC s.e.	Est.	Asy. s.e.	MC s.e.
Constant	-1.795***	0.567	0.0075	-1.752***	0.526	0.1015
AVGCA	-0.072***	0.018	0.0006	-0.074***	0.017	0.0028
AVGGROW	0.007	0.024	0.0002	0.007	0.026	0.0009
AVGINV	0.004	0.013	0.0002	0.004	0.015	0.0012
AVGTT	-0.161*	0.093	0.0027	-0.171*	0.094	0.0170
GOV	0.018	0.014	0.0001	0.018	0.015	0.0006
OT	-0.010	0.012	> 0.0001	-0.010	0.013	0.0007
OPEN	-0.108	0.109	0.0025	-0.116	0.118	0.0164
USINT	0.097	0.075	0.0007	0.098	0.082	0.0074
GROWOECD	0.057	0.087	0.0009	0.054	0.093	0.0070
INTPAY	0.029	0.035	0.0007	0.030	0.038	0.0037
RES	-0.097**	0.046	0.0016	-0.103**	0.047	0.0100
CONCDEB	-0.190**	0.088	0.0022	-0.199**	0.093	0.0152
κ	0.520	0.297	0.0172	0.486*	0.259	0.0764
σ_τ	0.142	0.322	0.0049	0.078*	0.051	1.9035
ρ	0.349*	0.198	0.0363	0.376**	0.175	0.0615
σ_e	1.077			1.082		
Log-likelihood	-253.27		0.0363	-253.20		0.3356

Note: The estimated model is given by Equation (3.2) and (3.4). The ML-EIS and ML-GHK estimation are based on a MC sample size of $S = 100$. The EIS simulator is based on three EIS iterations. The asymptotic standard errors are calculated as the square root of the diagonal elements of the inverse Hessian and the MC standard errors from 30 replications of the ML-EIS and ML-GHK estimation. *, **, and *** indicates statistical significance at the 10%, 5% and 1% significance level.

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

Table 3.5: ML-EIS estimates of Model 4: AR(1) time-specific effects

Variable	Estimate	Marg. Eff.
Constant	-1.967*** (0.677)	
AVGCA	-0.064*** (0.014)	-0.009
AVGGROW	0.013 (0.022)	0.002
AVGINV	-0.001 (0.011)	-0.0001
AVGTT	-0.122 (0.075)	-0.017
GOV	0.018 (0.012)	0.003
OT	-0.010 (0.011)	-0.001
OPEN	-0.065 (0.095)	-0.009
USINT	0.070 (0.071)	0.010
GROWOECD	0.113 (0.097)	0.016
INTPAY	0.011 (0.032)	0.002
RES	-0.073** (0.035)	-0.010
CONCDEB	-0.163** (0.074)	-0.023
κ	1.013*** (0.139)	0.210
σ_τ	0.154 (0.201)	
δ	-0.888*** (0.041)	
σ_ξ	0.089** (0.048)	
σ_e	1.030	
Log-likelihood	-253.13	
F -statistic for exogeneity of x_{it}	2.09	
t -statistic for exogeneity of y_{i0}	-1.98	

Note: The estimated model is given by Equations (3.2), (3.5), and (3.6). The ML-EIS estimation are based on a MC sample size of $S = 100$. The asymptotic standard errors are given in parentheses and obtained from the inverse Hessian. *, **, and *** indicates statistical significance at the 10%, 5% and 1% significance level. The 1% and 5% critical values of the F -statistic (t -statistic) are 2.71 and 2.03 (2.68 and 2.01).

seems to be able to separate the two sources of persistence. Also, the estimated autocorrelation parameter of -0.89 implies a strong mean reversion in the common time-specific factor. This mean-reverting tendency in the common factor affects the common probability of experiencing a current account reversal across all countries and is, therefore, fully consistent with a global accounting restriction requiring that deficits and surpluses across all national current accounts need to be balanced. In particular, one would expect that a temporary simultaneous increase in the propensities to experience a large reduction in current account deficits is immediately reverted in order to guarantee a global balance in deficits and surpluses, rather than a persistent and long-lasting increase in individual propensities. Although the time-specific factor capturing global shocks and/or contagion effects is significant, it appears to be quantitatively fairly small. In fact, the fraction of error variance due to the time-specific effect is only 3.5%. Therefore, it is not surprising that the overall fit of the model and its predicted dynamic effects (see, the lower right panel of Figure 3.1) do not change significantly relative to the pure random country-specific effect model in Table 3.2 which leaves out the time-specific effect.

Finally, we note that the quantitatively low impact of the common time-specific factor might be due to the implicit restriction that the loading w.r.t. that factor is the same across all countries. Hence, a natural extension of the model would be to allow for factor loadings, which differ across countries (whether randomly or deterministically). However, due to a substantial increase in the number of parameters or the dimension of the integration problem associated with the likelihood evaluation the statistical inference of such an extension is non-trivial without further restrictions and is left to future research.

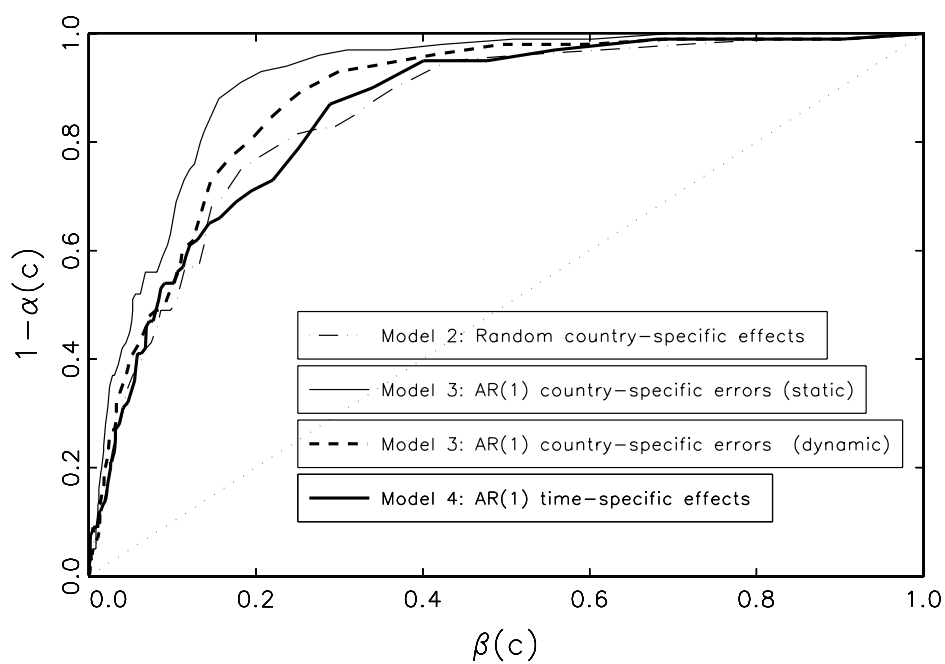
3.5.6 Predictive Performance

Models 2 to 4 are essentially observationally equivalent with log-likelihood values ranging from -253.1 to -255.2. However, log-likelihood comparisons provide an incomplete picture of the overall quality of a binary model. Hence, next models 2 to 4 are compared on two predictive benchmarks: the proportion of correctly predicted binary outcomes and predicted duration distribution of reversal episodes.

Assessing the predictive performance of an estimated binary model requires selecting a threshold c whereby success (current account reversal) is predicted

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

Figure 3.2: ROC Curve



Receiver Operating Characteristic for Models 2 and 4.

iff the predicted probability is larger than c , i.e., $r_{it} = \hat{p}(y_{it}|x_{it}, y_{it-1}) > c$. The corresponding classification error probabilities are given by

$$\alpha(c) = 1 - p(r_{it} > c|y_{it} = 1) \quad \text{and} \quad \beta(c) = p(r_{it} > c|y_{it} = 0), \quad (3.43)$$

which can be approximated by the corresponding relative frequencies of misclassification. Since the sample portion Π of success is only of the order of 0.1, it does not make sense to select a threshold c which minimizes the unconditional probability of misclassification $p(c) = \Pi\alpha(c) + (1 - \Pi)\beta(c)$. Following Winkelmann and Boes (2006), the threshold c_* which minimizes the sum of classification error probabilities $\alpha(c) + \beta(c)$ was computed for each model. Their Receiver Operating Characteristic (ROC) curves were also computed, which are defined as the curves plotting $1 - \alpha(c)$ against $\beta(c)$, as well as the areas under these ROC curves. These areas have a minimum of 0.5 (complete randomness) and a maximum of 1 (errorless classification). The ROC curves are displayed in Figure 3.2 and associated results for the optimal threshold c_* , classification error probabilities for c_* and ROC areas are reported in Table 6.

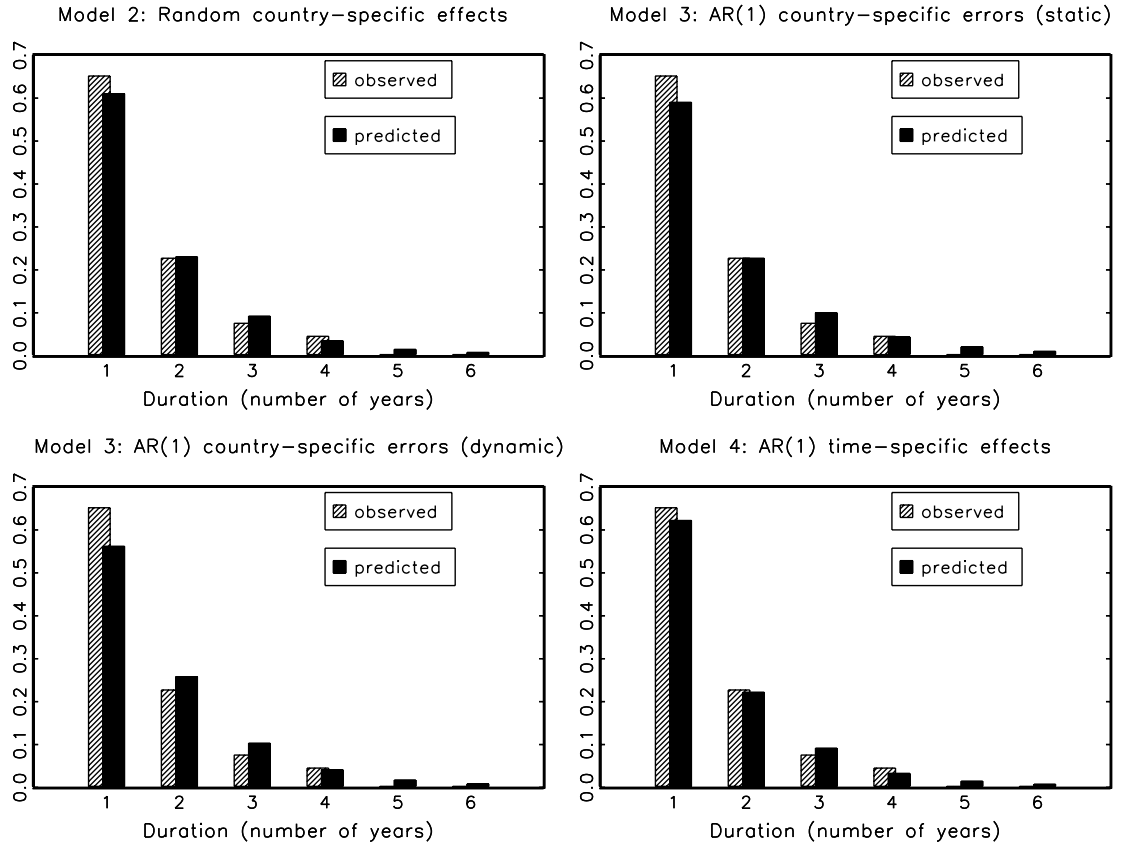
Note that c_* ranges from 0.08 to 0.11, which are close to the sample proportion Π of 0.10. Model 3 with AR(1) country-specific errors without state-dependence has the best predictive performance with $\alpha(c_*) + \beta(c_*) = 0.27$ and a ROC area of 0.91 (the corresponding figures for the other models range from 0.36 to 0.43 and 0.85 to 0.88, respectively). Also its ROC curve dominates those of the other models. Based on the optimal threshold it correctly predicts 91% of the observed reversals and 82% of the non-reversals.

Each estimated model was also used to simulate 20,000 fictitious panel data sets of the binary outcome conditional on the observed x_{it} variables in order to obtain accurate MC approximations of the predictive distributions of the duration of reversal episodes to be compared with the frequency distribution observed for the data (see Figure 3.3, and Table 3.6 for predicted average durations). It appears that models 2 and 4 have a better performance than model 3 with a better fit to the empirical distribution and predicted average durations closer to the observed average of 1.52. However, the differences across the models seem to be not large enough to overturn the ROC ranking. Thus, if the likelihood criterion, which by itself is fairly uninformative about the source of serial dependence, is supplemented by measures of predictive performance, the model with AR(1)

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

country-specific shocks and without state-dependence appears to be the preferred specification.

Figure 3.3: Duration of Reversal Episodes



Observed and predictive relative frequencies for the duration of reversal episodes for Models 2 to 4. The observed duration is 1.52 years.

3.6 Conclusion

This chapter uses different non-linear panel data specifications in order to investigate the causes and dynamics of current account reversals in low- and middle-income countries. In particular, four sources of serial persistence were analyzed: (i) a country-specific random effect reflecting time-invariant differences in institutional, political or economic factors; (ii) serially correlated transitory error

3.6 Conclusion

Table 3.6: Classification errors and predicted average duration in years

	c_*	$\alpha(c_*)$	$\beta(c_*)$	ROC area	average duration
Model 2: Random country-specific effects	0.11	0.25	0.18	0.85	1.68 (0.12)
Model 3: AR(1) country-specific errors (static)	0.12	0.09	0.18	0.91	1.77 (0.14)
Model 3: AR(1) country-specific errors (dynamic)	0.09	0.11	0.25	0.88	1.80 (0.14)
Model 4: AR(1) time-specific effects	0.08	0.13	0.28	0.86	1.66 (0.12)

Note: Estimated standard deviation of the predicted average duration are given in parentheses. The observed average duration is 1.52 years.

component capturing persistent country-specific shocks; (iii) dynamic common time-specific factor effects, designed to account for potential spill-over effects and global shocks to all countries; and (iv) a state dependence component to control for the effect of previous events of current account reversal and to capture slow adjustments in international trade flows.

The likelihood evaluation of panel models with country-specific random heterogeneity require univariate integrals that were efficiently solved using Gauss-Hermite quadrature. For likelihood-based estimation of panel models with country-specific random heterogeneity and serially correlated error components, the MC integration technique of Efficient Importance Sampling (EIS) was used. The application of EIS allows for numerically very accurate and reliable ML estimation of those models. In particular, it improves significantly the numerical efficiency of GHK, which is the most frequently used MC procedure to estimate non-linear panel models with serially correlated errors.

The empirical results indicate that the static pooled probit model is strongly dominated by the alternative models with serial dependence. However, state-dependence and transitory country-specific errors are essentially observationally equivalent, although the ROC curve gives more support to the model with transi-

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

tory country-specific errors and with no state-dependence. Autocorrelated errors and state-dependence are both significant only with the inclusion of random time-specific effects into the model together with state-dependence, even though the time-specific effect is small with limited effect on the overall fit of the model. Also, conclusive evidence for the existence of random country-specific effects was not found.

Overall, the results relative to the determinants of current account reversals are in line with the those in the empirical literature on current account crises and confirm the empirical relevance of theoretical solvency and sustainability considerations w.r.t. a country's trade balance. In particular, countries with high current account imbalances, low foreign reserves, a small fraction of concessional debt, and unfavorable terms of trades are more likely to experience a current account reversal. These results are fairly robust against the dynamic specification of the model.

3.A1 Appendix 1: EIS-implementation

This appendix details the implementation of the EIS procedure for the panel probit model (3.2) and (3.4) to obtain MC estimates for the likelihood contribution $I(\theta)$ given by equation (3.10). In order to take full advantage of the properties of Gaussian kernels, we adopt the following conventions:

(i) Gaussian kernels are represented under their natural parametrization – see Lehmann (1986, Section 2.7). Whence, $k_t(\underline{\lambda}_t; a_t)$ in Equation (3.26) is parameterized as

$$k_t(\underline{\lambda}_t; a_t) = \frac{\mathbb{I}(\epsilon_t \in D_t^*)}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} (\underline{\lambda}_t' P_t \underline{\lambda}_t + 2 \underline{\lambda}_t' q_t) \right\}, \quad (3.A1-1)$$

with $D_t^* = (-\infty, \gamma_t + \delta_t \lambda_0]$. The EIS parameter a_t consists of the six lower diagonal elements of P_t and the three elements in q_t (the positivity constraint on P_t never binds in our application).

(ii) All factorizations of k_t are based upon the Cholesky decomposition of P_t into

$$P_t = L_t \Delta_t L_t', \quad (3.A1-2)$$

where $L_t = \{l_{ij,t}\}$ is a lower triangular matrix with ones on the diagonal and Δ_t is diagonal matrix with diagonal elements $d_{i,t} \geq 0$. Let

$$l_{1,t} = (l_{21,t}, l_{31,t})', \quad l_{2,t} = (1, l_{32,t})'. \quad (3.A1-3)$$

The key steps in our EIS implementation consists of finding the analytical expression of $\chi_t(\underline{\eta}_{t-1}; a_t)$. It is the object of the following lemma.

Lemma 1. *The integral of $k_t(\underline{\lambda}_t; a_t)$, as defined in (3.A1-1), w.r.t. ϵ_t is of the form given by Equation (3.29) together with*

$$k_{2,t}(\underline{\eta}_{t-1}; \cdot) = \exp \left\{ -\frac{1}{2} (d_{2,t} \eta_{t-1}' l_{2,t} l_{2,t}' \eta_{t-1} + 2 \eta_{t-1}' l_{2,t} m_{2,t}) \right\} \quad (3.A1-4)$$

$$k_{3,t}(\lambda_0; \cdot) = \exp \left\{ -\frac{1}{2} (d_{3,t} \lambda_0^2 + 2 m_{3,t} \lambda_0) \right\} \cdot r_t \quad (3.A1-5)$$

$$\alpha_t = \sqrt{d_{1,t}} \left(\gamma_t + \frac{m_{1,t}}{d_{1,t}} \right), \quad \beta_t = \sqrt{d_{1,t}} (l_{1,t} + \delta_t l) \quad (3.A1-6)$$

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

$$r_t = \frac{1}{\sqrt{d_{1,t}}} \exp \left\{ \frac{1}{2} \frac{m_{1,t}^2}{d_{1,t}} \right\}, \quad (3.A1-7)$$

with

$$m_t = \{m_{i,t}\} = L_t^{-1} q_t, \quad l' = (0, 1). \quad (3.A1-8)$$

Proof. The proof is straightforward under the Cholesky factorization introduced in (3.A1-2), deleting the index t for the ease of notation. First the transformation $z = L'\underline{\lambda}$ is introduced, whereby

$$z_1 = \epsilon + l'_1 \underline{\eta}_{-1}, \quad z_2 = l'_2 \underline{\eta}_{-1}, \quad z_3 = \lambda_0.$$

Whence,

$$\begin{aligned} \chi(\underline{\eta}_{-1}; \cdot) &= \exp \left\{ -\frac{1}{2} \left[\sum_{i=2}^3 (d_i z_i^2 + 2m_i z_i) \right] \right\} \\ &\quad \times \frac{1}{\sqrt{2\pi}} \int_{D_t^{**}} \exp \left\{ -\frac{1}{2} (d_1 z_1^2 + 2m_1 z_1) \right\} dz_1, \end{aligned}$$

where $D_t^{**} = (-\infty, \gamma_t + (l_{1,t} + \delta_t l')' \underline{\eta}_{-1}]$. Next, the quadratic form in z_1 is completed under the integral sign and following transformation is introduced

$$v = \sqrt{d_1} \left(z_1 + \frac{m_1}{d_1} \right).$$

The result immediately follows. \square

Next, full details of the recursive EIS implementation in Equations (3.19) to (3.24) are provided.

• *Period $t = T$:* With $\chi_{T+1} \equiv 1$, the only component of k_T is φ_T itself. Whence,

$$P_T = e_\rho e'_\rho \quad \text{and} \quad q_T = 0, \quad (3.A1-9)$$

with

$$e'_\rho = (1, -\rho, 0).$$

• *Period t ($T > t > 1$):* Given Equation (3.29) together with lemma 1, the product $\varphi_t \cdot \chi_{t+1}$ comprises the following factors: φ_t as defined in Equation (3.11), $k_{2,t+1}$ as given by Equation (3.A1-4) and $\Phi(\alpha_{t+1} + \beta'_{t+1} \underline{\eta}_t)$, where $(\alpha_{t+1}, \beta_{t+1})$ are defined in Equation (3.A1-6). The first two factors are already Gaussian kernels.

3.A1 Appendix 1: EIS-implementation

Furthermore, the term $\Phi(\cdot)$ depends on $\underline{\lambda}_t$ only through the linear combination $\beta'_{t+1}\underline{\eta}_t$. Whence, $k_{0,t}$ in Equation (3.25) is defined as

$$k_{0,t}(\underline{\lambda}_t; a_t) = k_{2,t+1}(\underline{\eta}_t, \cdot) \exp \left\{ -\frac{1}{2} \left[a_{1,t}(\beta'_{t+1}\underline{\eta}_t)^2 + 2a_{2,t}(\beta'_{t+1}\underline{\eta}_t) \right] \right\}, \quad (3.A1-10)$$

with $a_t = (a_{1,t}, a_{2,t})$. It follows that $k_{2,t+1}$ also cancels out in the auxiliary EIS regressions which simplifies into OLS of $\ln \Phi(\alpha_{t+1} + \beta'_{t+1}\underline{\eta}_t)$ on $\beta'_{t+1}\underline{\eta}_t$ and $(\beta'_{t+1}\underline{\eta}_t)^2$ together with a constant. From this EIS regressions one obtains estimated values for $(a_{1,t}, a_{2,t})$. Note that $\underline{\eta}_t$ can be written as

$$\underline{\eta}_t = A\underline{\lambda}_t, \quad \text{with} \quad A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (3.A1-11)$$

It follows that the parameters of the EIS kernel k_t in Equation (3.A1-1) are given by

$$P_t = e_\rho e'_\rho + d_{2,t+1} A' l_{2,t+1} l'_{2,t+1} A + a_{1,t} A' \beta_{t+1} \beta'_{t+1} A \quad (3.A1-12)$$

$$q_t = A' l_{2,t+1} m_{2,t+1} + a_{2,t} A' \beta_{t+1}. \quad (3.A1-13)$$

Its integrating factor $\chi_t(\underline{\eta}_t; a_t)$ follows by application of lemma 1.

· *Period $t = 1$* : The same principle as above applies to period 1, but requires adjustments in order to account for the initial condition. Specifically, we have

$$\underline{\lambda}_1 = \underline{\eta}_1 = (\epsilon_1, \lambda_0)', \quad \lambda_0 = \eta_0 (= \tau). \quad (3.A1-14)$$

This amounts to replacing A by I_2 in Equations (3.A1-11) to (3.A1-13). Whence, the kernel $k_1(\lambda_1, a_1)$ needs only be bivariate with

$$P_1 = e_1 e'_1 + d_{2,2} l_{2,2} l'_{2,2} + \hat{a}_{1,1} \beta_2 \beta'_2 \quad (3.A1-15)$$

$$q_1 = l_{2,2} m_{2,2} + \hat{a}_{2,1} \beta_2, \quad (3.A1-16)$$

with $e'_1 = (1, 0)$. Essentially, P_1 and q_1 have lost their middle row and/or column. To avoid changing notation in lemma 1, the Cholesky decomposition of P_1 is parameterized as

$$L_1 = \begin{pmatrix} 1 & 0 \\ l_{31,1} & 1 \end{pmatrix}, \quad D_1 = \begin{pmatrix} d_{1,1} & 0 \\ 0 & d_{3,1} \end{pmatrix}, \quad l_{1,1} = l_{31,1}, \quad (3.A1-17)$$

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

while $d_{2,2}$ and $l_{2,2}$ are now zero. Under these adjustments in notation, lemma 1 still applies with $k_2(\eta_0; \cdot) \equiv 1$ and β_1 reduced to the scalar

$$\beta_1 = \sqrt{d_{1,1}}(l_{1,1} + \delta_1). \quad (3.A1-18)$$

• *Period* $t = 0$ (untruncated integral w.r.t. $\lambda_0 \equiv \tau$): Accounting for the back transfer of $\{k_{3,t}(\lambda_0; \cdot)\}_{t=1}^T$, all of which are Gaussian kernels, the λ_0 -kernel is given by

$$k_0(\lambda_0; \cdot) = f_\tau(\lambda_0) \cdot \prod_{t=1}^T k_{3,t}(\lambda_0; \cdot) \cdot \exp \left\{ -\frac{1}{2} (\hat{a}_{1,0}\lambda_0^2 + 2\hat{a}_{2,0}\lambda_0) \right\}, \quad (3.A1-19)$$

where $(\hat{a}_{1,0}, \hat{a}_{2,0})$ are the (fixed point) coefficients of the EIS approximation of $\ln \Phi(\alpha_1 + \beta_1 \lambda_0)$. Note that k_0 is the product of $T + 2$ Gaussian kernels in λ_0 and is, therefore, itself a Gaussian kernel, whose mean m_0 and variance v_0^2 trivially obtain by addition from Equation (3.A1-19).

As mentioned above, fixed point convergence of the EIS auxiliary regressions (3.24) as well as continuity of corresponding likelihood estimates require the use of CRNs. In order to draw the ϵ_t 's from their (truncated) Gaussian samplers $m_t(\epsilon_t | \tilde{\eta}_{t-1}; \hat{a}_t)$ based on CRNs one can use the following result: If ϵ follows a truncated $N(\mu, \sigma^2)$ distribution with $b_l < \epsilon < b_u$, then a ϵ -draw is obtained as (see, e.g., Train, 2003)

$$\tilde{\epsilon} = \mu + \sigma \Phi^{-1} \left[\Phi \left(\frac{b_l - \mu}{\sigma} \right) + \tilde{u} \cdot \left\{ \Phi \left(\frac{b_u - \mu}{\sigma} \right) - \Phi \left(\frac{b_l - \mu}{\sigma} \right) \right\} \right], \quad (3.A1-20)$$

where \tilde{u} is a canonical draw from the $U(0, 1)$ distribution which is independent from the parameters indexing the distribution of ϵ .

3.A2 Appendix 2: MC experiments on the numerical efficiency of EIS

In order to compare the numerical accuracy of EIS relative to GHK, the integral in Equation (3.10) is evaluated under 18 artificial data sets considering three different values of ρ (0.2, 0.5, 0.9), two different values of σ_τ (0.1, 0.5) and three different sample sizes ($T=10, 20, 50$). All other coefficients are set equal to zero.

For each of these eighteen data sets, 200 i.i.d MC estimates of I_i based upon different sets of CRNs are produced. Both GHK and EIS individual estimates are

3.A2 Appendix 2: MC experiments on the numerical efficiency of EIS

based upon $S = 200$ auxiliary draws and the number of EIS iterations is fixed at three. In Table A1 we report the means, MC standard deviations and coefficients of variation of the 200 replications under both methods for the eighteen scenarios. As discussed in Richard and Zhang (2007), these standard deviations provide direct measures of numerical accuracy. Note immediately that the MC standard deviation under EIS are systematically lower than those under GHK, by factors ranging from 7 to 315.

As expected numerical accuracy is a decreasing function of ρ , σ_τ and T . This is especially the case for GHK with coefficient of variations above 1 for $T = 50$ and $\rho = 0.9$. In sharp contrast, the worse case scenario for EIS ($T = 50$, $\rho = 0.9$, $\sigma_\tau = 0.5$) has a coefficient of variation of 0.017. A more extensive and detailed comparison of GHK and EIS can be found in Liesenfeld and Richard (2007).

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

Table 3.7: List of countries

Country	Initial Obs.	Final Obs.	Reversals
Argentina	1988	2001	2
Bangladesh	1984	2000	0
Benin	1984	1999	2
Bolivia	1984	2001	3
Botswana	1984	2000	2
Brazil	1984	2001	1
Burkina Faso	1984	1992	0
Burundi	1989	2001	0
Cameroon	1984	1993	0
Central African Republic	1984	1992	0
Chile	1984	2001	3
China	1986	2001	1
Colombia	1984	2001	4
Congo Rep.	1984	2001	2
Costa Rica	1984	2001	1
Cote d'Ivoire	1984	2001	5
Dominican Republic	1984	2001	2
Ecuador	1984	2001	1
Egypt	1984	2001	3
El Salvador	1984	2001	2
Gabon	1984	1997	3
Gambia	1984	1995	1
Ghana	1984	2001	1
Guatemala	1984	2001	0
Guinea-Bissau	1987	1995	0
Haiti	1984	1998	3
Honduras	1984	2001	2
Hungary	1986	2001	1
India	1984	2001	0
Indonesia	1985	2001	1

3.A2 Appendix 2: MC experiments on the numerical efficiency of EIS

Table 3.8: List of countries

Country	Initial Obs.	Final Obs.	Reversals
Jordan	1984	2001	4
Kenya	1984	2001	2
Lesotho	1984	2000	0
Madagascar	1984	2001	0
Malawi	1984	2001	0
Malaysia	1984	2001	5
Mali	1991	2000	0
Mauritania	1984	1996	4
Mexico	1984	2001	1
Morocco	1984	2001	2
Niger	1984	1993	1
Nigeria	1984	1997	2
Pakistan	1984	2001	3
Panama	1984	2001	2
Paraguay	1984	2001	2
Peru	1984	2001	2
Philippines	1984	2001	3
Rwanda	1984	2001	1
Senegal	1984	2001	3
Seychelles	1989	2001	4
Sierra Leone	1984	1995	0
Sri Lanka	1984	1997	2
Swaziland	1984	2001	3
Thailand	1984	2001	3
Togo	1984	2000	0
Tunisia	1984	2001	2
Turkey	1984	2001	0
Uruguay	1984	2001	0
Venezuela	1984	2001	1
Zimbabwe	1984	1992	2

3. ESTIMATION OF DYNAMIC PANEL PROBIT MODELS

Table A1. Monte Carlo simulation results of GHK and EIS

		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.9$	
		EIS	GHK	EIS	GHK	EIS	GHK
$T = 10$							
$\sigma_\tau = 0.1$	mean	9.8e-04	9.8e-04	1.9e-03	1.9e-03	4.7e-03	4.7e-03
	std. dev.	1.7e-07	1.9e-05	4.5e-06	1.0e-04	1.9e-05	5.3e-04
	coeff. var.	1.7e-04	2.0e-02	2.4e-03	5.5e-02	4.2e-03	1.1e-01
$\sigma_\tau = 0.5$	mean	7.9e-04	7.9e-04	5.5e-03	5.4e-03	1.4e-01	1.4e-01
	std. dev.	5.4e-07	3.6e-05	2.4e-05	3.4e-04	1.5e-03	9.8e-03
	coeff. var.	6.9e-04	4.5e-02	4.3e-03	6.3e-02	1.1e-02	7.0e-02
$T = 20$							
$\sigma_\tau = 0.1$	mean	6.9e-07	6.9e-07	1.0e-06	9.7e-07	6.7e-06	5.9e-06
	std. dev.	1.4e-10	2.2e-08	3.1e-09	8.8e-08	7.7e-08	1.5e-06
	coeff. var.	2.0e-04	3.1e-02	3.0e-03	9.0e-02	1.2e-02	2.6e-01
$\sigma_\tau = 0.5$	mean	7.8e-07	7.7e-07	2.6e-06	2.5e-06	8.3e-06	6.6e-06
	std. dev.	6.5e-10	4.6e-08	1.5e-08	2.7e-07	1.1e-07	1.6e-06
	coeff. var.	8.3e-04	6.0e-02	6.0e-03	1.0e-01	1.4e-02	2.4e-01
$T = 50$							
$\sigma_\tau = 0.1$	mean	1.4e-15	1.4e-15	1.1e-14	1.0e-14	7.6e-10	8.3e-10
	std. dev.	2.9e-19	9.2e-17	3.9e-17	2.3e-15	1.1e-11	9.7e-10
	coeff. var.	2.0e-04	6.4e-02	3.5e-03	2.2e-01	1.5e-02	1.2e+00
$\sigma_\tau = 0.5$	mean	3.5e-15	3.5e-15	4.8e-15	4.8e-15	1.1e-09	1.3e-09
	std. dev.	2.3e-18	4.0e-16	1.8e-17	1.6e-15	1.8e-11	1.8e-09
	coeff. var.	6.6e-04	1.1e-01	3.7e-03	3.4e-01	1.7e-02	1.4e+00

Note: MC-estimation of the likelihood contribution I_i under the panel model (3.2) and (3.4) for a simulated fictitious sample $\{y_{it}\}_{t=1}^T$ (see Equation 3.10). The mean, the standard deviation and the coefficient of variation are obtained from 200 independent replications of the MC estimation of I_i . The EIS and GHK MC-estimates are based upon a simulation sample size $S = 200$.

Chapter 4

Nonlinear State-Space Models

Filtering is the problem of estimating the state of a stochastic dynamic system from noisy measurement observations (Jazwinski, 1970). It dates back to more than two centuries, when Gauss was interested in determining the orbital elements of a celestial body based on many measurements over time, and developed the *least squares* technique. Filtering is important in the time series analysis of state-space models not only as a means to infer the state of the system, but also as a means to compute the likelihood function and to estimate the parameters of such models (see Durbin and Koopman, 2001; Harvey, 1990).

Likelihood evaluation and filtering in applications involving state-space models require 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, either via deterministic quadrature methods or via MC integration. Since deterministic methods become unfeasible when the number of states is large, sequential Monte Carlo methods have become the standard tool in nonlinear filtering (see Doucet et al., 2001).

The most widely used sequential MC filter is the particle filter developed by Gordon et al. (1993) and Kitagawa (1996). 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, and examples of its use are becoming widespread; in economics, e.g., see Kim et al. (1998) for an application involving stochastic volatility models; and Fernandez-Villaverde

4. NONLINEAR STATE-SPACE MODELS

and Rubio-Ramirez (2005, 2007) for applications involving dynamic stochastic general equilibrium (DSGE) 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 (see Pitt, 2002). Second, the supports upon which approximations are based are not adapted: the support of period- t importance sampler used to approximate the integrals in the filtering process only incorporate information conveyed by values of the observable variables available in period $t - 1$, but not period t (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 et al. (2001); Pitt (2002); Ristic et al. (2004), 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 importance samplers via the use of information available through period t . However, with the exception of the extension proposed by Pitt (2002), who employs a bootstrap-smoothing approximation designed to address this problem for the specialized case in which the state space is unidimensional, 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.

This chapter, which was co-authored with David DeJong, Dharmarajan Hariharan, Roman Liesenfeld, and Jean-François Richard, presents the EIS filter, a sequential Monte Carlo filter based on adapted period- t importance samplers, 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) (see section 2.2.3). Construction is facilitated using an optimization procedure designed to

4.1 Likelihood Evaluation and Filtering in State-Space Representations

minimize numerical standard errors associated with the approximated integral. Example applications involve the analysis of DSGE models, and are used to illustrate the relative performance of the particle and EIS filters.

A brief literature review is helpful to motivate the focus on the analysis of DSGE models. The pioneering work of Sargent (1989) demonstrated the mapping of DSGE models into linear/Gaussian state-space representations amenable to likelihood-based analysis achievable via the Kalman filter. DeJong et al. (2000) developed a Bayesian approach to analyzing these models. Subsequent work has involved the implementation of DSGE models towards a broad range of empirical objectives, including forecasting and guidance of the conduct of aggregate fiscal and monetary policy (see Smets and Wouters, 2003).

Prior to the work of Fernandez-Villaverde and Rubio-Ramirez (2005, 2007), likelihood-based implementation of DSGE models was conducted using linear and Gaussian representations. But their findings revealed an important caveat: approximation errors associated with linear representations of DSGE models can impart significant errors in corresponding likelihood representations. As a remedy, they demonstrated use of the particle filter for achieving likelihood evaluation for non-linear model representations. But as the examples in this chapter illustrate, the numerical inefficiencies noted above suffered by the particle filter can be acute in applications involving DSGE models. By eliminating these inefficiencies, the EIS filter offers a significant advance in the empirical analysis of DSGE models.

4.1 Likelihood Evaluation and Filtering in State-Space Representations

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 . State-space representations consist of a state-transition equation

$$s_t = \gamma(s_{t-1}, Y_{t-1}, v_t), \quad (1)$$

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

$$y_t = \delta(s_t, Y_{t-1}, u_t), \quad (2)$$

4. NONLINEAR STATE-SPACE MODELS

where u_t is a vector of innovations with respect to (s_t, Y_{t-1}) . Hereafter, v_t are referred to as structural shocks, and u_t as measurement errors.

Given a realization of the sequence of observations Y_t , the state estimation problem consists of computing an estimate of s_k based on Y_t . If $k < t$, this problem is referred as smoothing, while if $k = t$ the problem is called filtering, and $k > t$ defines the prediction problem. Filtering and prediction are usually associated with real-time operations, in which estimates are required given the observations available now. If inference on some past states of the system are required, smoothing makes use of present information to improved the past states estimates.

The likelihood function $f(Y_T)$ is obtained by interpreting (1) and (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. Since the representation is recursive, $f(Y_T)$ factors sequentially as

$$f(Y_T) = \prod_{t=1}^T f(y_t|Y_{t-1}), \quad (3)$$

where $f(y_1|Y_0) \equiv f(y_1)$. The time- t likelihood $f(y_t|Y_{t-1})$ is obtained by marginalizing over s_t :

$$f(y_t|Y_{t-1}) = \int f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1}) ds_t, \quad (4)$$

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

$$f(s_t|Y_{t-1}) = \int f(s_t|s_{t-1}, Y_{t-1}) f(s_{t-1}|Y_{t-1}) ds_{t-1}, \quad (5)$$

and $f(s_{t-1}|Y_{t-1})$ is the time- $(t-1)$ filtering density. Advancing the time subscript by one period, from Bayes' theorem, $f(s_t|Y_t)$ is given by

$$f(s_t|Y_t) = \frac{f(y_t, s_t|Y_{t-1})}{f(y_t|Y_{t-1})} = \frac{f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1})}{f(y_t|Y_{t-1})}. \quad (6)$$

Likelihood construction is achieved by calculating (4) and (5) sequentially from periods 1 to T , taking as an input in period t the filtering density constructed in period $(t-1)$. In period 1 the filtering density is the known marginal density $f(s_0)$, which can be degenerate as a special case; i.e., $f(s_0|Y_0) \equiv f(s_0)$.

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 (6) and (4), this can be written as

$$E_t(h(s_t)|Y_t) = \frac{\int h(s_t) f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1}) ds_t}{\int f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1}) ds_t}. \quad (7)$$

4.2 The Particle Filter and Leading Extensions

Since the proposed procedure is an extension of the particle filter developed by Gordon et al. (1993) and Kitagawa (1996), this Section provides a brief overview of it. 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 (5). This is done by drawing $s_t^{1,i}$ from the conditional density $f(s_t|s_{t-1}^{0,i}, Y_{t-1})$, $i = 1, \dots, N$. 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 (4) is given by

$$\widehat{f}_N(y_t|Y_{t-1}) = \frac{1}{N} \sum_{i=1}^N f(y_t|s_t^{1,i}, Y_{t-1}). \quad (8)$$

Next, $f(s_t|Y_t)$ is approximated by re-weighting $\{s_t^{1,i}\}_{i=1}^N$ in accordance with (6) (the updating step): a particle $s_t^{1,i}$ with prior weight $\frac{1}{N}$ is assigned the posterior weight

$$w_t^{0,i} = \frac{f(y_t|s_t^{1,i}, Y_{t-1})}{\sum_{j=1}^N f(y_t|s_t^{1,j}, Y_{t-1})}. \quad (9)$$

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 strengths and weaknesses, which are well documented in previous studies, can be pinpointed. Its strength lies in its simplicity: the algorithm described above is straightforward and universally applicable.

Its weaknesses are twofold. 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

4. NONLINEAR STATE-SPACE MODELS

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).

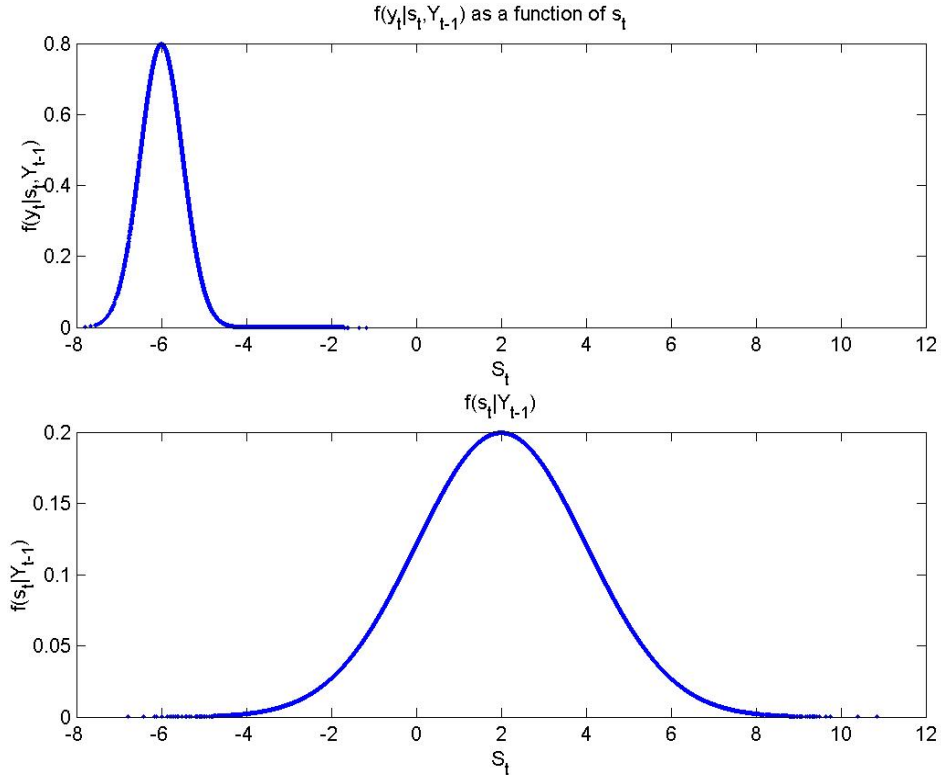
Second, as the filter enters period t , the discrete approximation of $f(s_{t-1}|Y_{t-1})$ is set. Hence the swarm $\{s_t^{1,i}\}_{i=1}^N$ produced in the augmentation stage ignores information provided by y_t , thus Pitt and Shephard (1999) refer to these augmenting draws as “blind”. It follows that 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 $f(y_t|s_t, Y_{t-1})$. The lower panel of Figure 4.1) represents $f(s_t|Y_{t-1})$, the sampler used in the particle filter to solve the likelihood integral (4), while the upper panel shows the measurement density $f(y_t|s_t, Y_{t-1})$ as a function of s_t in a situation where it is peaked at the tails of $f(s_t|Y_{t-1})$. Note that most of the particles coming from this sampler will be placed far from the importance region defined by the product $f(y_t|s_t, Y_{t-1}) \cdot f(s_t|Y_{t-1})$. 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 et al. (1993) and Kitagawa (1996) consists simply of making $N' \gg N$ blind proposals $\{s_t^{1,j}\}_{j=1}^{N'}$ as with the particle filter, and then obtaining the swarm $\{s_t^{0,i}\}_{i=1}^N$ 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 et al. (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.

4.2 The Particle Filter and Leading Extensions

Figure 4.1: Sample Impoverishment



Pitt and Shephard (1999) developed an extension that is closely related to the EIS Filter. 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 (4), but with $f(s_t|Y_{t-1})$ unknown, importance sampling achieves approximation via the introduction into the integral of an importance density $g(s_t|Y_t)$:

$$f(y_t|Y_{t-1}) = \int \frac{f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1})}{g(s_t|Y_t)} g(s_t|Y_t) ds_t. \quad (10)$$

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

$$\hat{f}(y_t|Y_{t-1}) \approx \frac{1}{N} \sum_{i=1}^N \frac{f(y_t|s_t^{0,i}, Y_{t-1}) f(s_t^{0,i}|Y_{t-1})}{g(s_t^{0,i}|Y_t)}. \quad (11)$$

Pitt and Shephard 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 the

4. NONLINEAR STATE-SPACE MODELS

product $f(y_t|s_t, Y_{t-1}) f(s_t|Y_{t-1})$, rendering the ratios in (11) as constants. Pitt and Shephard viewed adaption as computationally infeasible, due to the requirement of computing $f(s_t^{0,i}|Y_{t-1})$ 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 $s_t = \mu_t^k = E(s_t|s_{t-1}^{0,k}, Y_{t-1})$. A zero-order expansion yields their auxiliary particle filter; a first-order expansion yields their adapted particle filter, and 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 $\{s_{t-1}^{0,i}\}$ 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 $\{s_{t-1}^{0,i}\}$. 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 (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 $\{s_t^{0,i}\}_{i=1}^N$. This enables the use of common random numbers (CRNs) to produce likelihood estimates that are continuous functions of model parameters (see Hendry, 1984, or Section 2.2.1).

4.3 Parametric EIS Filter

EIS is an automated procedure for constructing continuous importance samplers fully adapted as global approximations to targeted integrands as detailed in Section 2.2.3. In subsection 4.3.1 only the general principles behind EIS will be outlined, in the context of evaluating (4). Subsection 4.3.2 discusses the computation of $f(s_t|Y_{t-1})$ in (4) at auxiliary values of s_t generated under period- t EIS optimization. Subsection 4.3.3 discusses a special case that often characterizes state-space representations: degenerate transition densities.

4.3.1 EIS integration

Let $\varphi_t(s_t) = f(y_t|s_t, Y_{t-1})f(s_t|Y_{t-1})$ in (4), 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 m are

$$m(s_t; a_t) = \frac{k(s_t; a_t)}{\chi(a_t)}, \quad \chi(a_t) = \int k(s_t; a_t) ds_t. \quad (12)$$

The selection of K is problem-specific; here Gaussian specifications will be discussed; DeJong et al. (2008) discusses an extension to piecewise-continuous specifications. The objective of EIS is to select the parameter value $\hat{a}_t \in A$ that minimizes the variance of the ratio $\frac{\varphi_t(s_t)}{m(s_t|a_t)}$ over the range of integration. Following Richard and Zhang (2007), a (near) optimal value \hat{a}_t is obtained as the solution to

$$(\hat{a}_t, \hat{c}_t) = \arg \min_{a_t, c_t} \int [\ln \varphi_t(s_t) - c_t - \ln k(s_t; a_t)]^2 m(s_t; a_t) ds_t, \quad (13)$$

where c_t is an intercept meant to calibrate $\ln(\varphi_t/k)$. Equation (13) is a standard least squares problem, except that the auxiliary sampling density itself depends upon a_t . This is resolved by reinterpreting (13) as the search for a fixed-point solution. An operational MC version implemented (typically) using $R \ll N$ draws, is as follows:

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

$$(\hat{a}_t^{l+1}, \hat{c}_t^{l+1}) = \arg \min_{a_t, c_t} \sum_{i=1}^R [\ln \varphi_t(s_{t,l}^i) - c_t - \ln k(s_{t,l}^i; a_t)]^2. \quad (14)$$

If K belongs to the exponential family of distributions, there exists a parameterization a_t such that the auxiliary problems in (14) are linear, as explained in Section 2.2.3.1.

Three technical points bear mentioning here. First, the evaluation of $\varphi_t(s_t)$ entails the evaluation of $f(s_t|Y_{t-1})$, which is unavailable analytically and must be approximated; this is discussed below in Section 4.3.3. Second, the selection of the initial value \hat{a}_t^1 is important for achieving rapid convergence; Section 4.4 presents an effective algorithm for specifying \hat{a}_t^1 in applications involving DSGE models

4. NONLINEAR STATE-SPACE MODELS

(one step in each of the examples considered). Third, to achieve rapid 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 ; e.g., standardized Normal draws when m is Gaussian).

An additional substantive point also bears mentioning. At convergence to \widehat{a}_t , the EIS sampler $m(s_t; \widehat{a}_t)$ not only provides the optimal global approximation to the targeted integrand $\varphi_t(s_t) = f(y_t|s_t, Y_{t-1}) \cdot f(s_t|Y_{t-1})$, but also serves as the optimized approximation to the time- t filtering density $f(s_t|Y_t)$. Thus as with the particle filter, the EIS filter facilitates likelihood evaluation and filtering simultaneously.

At convergence to \widehat{a}_t , the EIS filter approximation of $f(y_t|Y_{t-1})$ in (4) is given by

$$\widehat{f}_N(y_t|Y_{t-1}) = \frac{1}{N} \sum_{i=1}^N \omega(s_t^i; \widehat{a}_t), \quad (15)$$

$$\omega(s_t^i; \widehat{a}_t) = \frac{f(y_t|s_t^i, Y_{t-1}) f(s_t^i|Y_{t-1})}{m(s_t^i; \widehat{a}_t)}, \quad (16)$$

where $\{s_t^i\}_{i=1}^N$ are drawn from the (final) EIS sampler $m(s_t; \widehat{a}_t)$. This estimate converges almost surely towards $f(y_t|Y_{t-1})$ under weak regularity conditions (see Geweke, 1989a, or Section 2.2.2). Violations of these conditions typically result from the use of samplers with thinner tails than those of φ_t . As mentioned in Section 2.2.3, Richard and Zhang (2007) offer a diagnostic measure that is adept at detecting this problem, since it 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, in order to generate draws in the tails of $m(s_t; \widehat{a}_t)$ and to examine their effect on the variance of the ratio $\frac{\varphi_t}{m}$.

Pseudo-code for implementing the EIS filter is as follows:

- At period t , the sampler $m(s_{t-1}; \widehat{a}_{t-1})$, and corresponding draws and weights $\{s_{t-1}^i, \omega(s_{t-1}^i; \widehat{a}_{t-1})\}_{i=1}^N$ are inherited from period $t-1$, where in period 0 $m(s_0; \widehat{a}_0) \equiv f(s_0)$.
- Using an initial value \widehat{a}_t^1 , obtain R draws $\{s_t^{i,l}\}_{i=1}^R$ from $m(s_t; \widehat{a}_t^1)$, and solve (14) to obtain \widehat{a}_t^2 . Repeat until convergence, yielding \widehat{a}_t .

- Obtain N values $\{s_t^i\}_{i=1}^N$ from the optimized sampling density $m(s_t; \hat{a}_t)$, and calculate (15).
- Pass $m(s_t; \hat{a}_t)$ and $\{s_t^i, \omega(s_t^i; \hat{a}_t)\}_{i=1}^N$ to period $t + 1$. Repeat until period T is reached.

As it will be explained below, $\{s_t^i, \omega(s_t^i; \hat{a}_t)\}_{i=1}^N$ are passed from period t to $t + 1$ to facilitate the approximation of the unknown $f(s_t|Y_{t-1})$ appearing in (14) and (16).

4.3.2 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 three operational alternatives for overcoming this hurdle will be discussed. Below, S denotes the number of points used for each individual evaluation of $f(s_t|Y_{t-1})$.

Weighted-sum approximations

Combining (5) and (6), it is possible to rewrite $f(s_t|Y_{t-1})$ as a ratio of integrals:

$$f(s_t|Y_{t-1}) = \frac{\int f(s_t|s_{t-1}, Y_{t-1})f(y_{t-1}|s_{t-1}, Y_{t-2})f(s_{t-1}|Y_{t-2}) ds_{t-1}}{\int f(y_{t-1}|s_{t-1}, Y_{t-2})f(s_{t-1}|Y_{t-2}) ds_{t-1}}, \quad (17)$$

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

$$\hat{f}_S(s_t|Y_{t-1}) = \frac{\sum_{i=1}^S f(s_t|s_{t-1}^i, Y_{t-1}) \cdot \omega(s_{t-1}^i; \hat{a}_{t-1})}{\sum_{i=1}^S \omega(s_{t-1}^i; \hat{a}_{t-1})}, \quad (18)$$

where $\{s_{t-1}^i\}_{i=1}^S$ denotes EIS draws from $m(s_{t-1}|\hat{a}_{t-1})$, and $\{\omega(s_{t-1}^i; \hat{a}_{t-1})\}_{i=1}^S$ denotes associated weights (both of which are carried over from period- $t - 1$).

Obviously $m(s_{t-1}|\hat{a}_{t-1})$ is not an EIS sampler for the numerator in (17). 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 $m(s_{t-1}|\hat{a}_{t-1})$. 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 the same set of draws for the numerator and the denominator are used typically creates positive correlation between their respective MC estimators, thus reducing the variance of their ratio (see Geweke, 1989a).

4. NONLINEAR STATE-SPACE MODELS

A constant weight approximation

When EIS delivers a close global approximation to $f(s_{t-1}|Y_{t-1})$, the weights $\omega(s_{t-1}; \hat{a}_{t-1})$ will be near constants over the range of integration. Replacing these weights by their arithmetic means $\bar{\omega}(\hat{a}_{t-1})$ in (17) and (18), the following simplification is obtained:

$$f(s_t|Y_{t-1}) \simeq \int f(s_t|s_{t-1}, Y_{t-1}) \cdot m(s_{t-1}; \hat{a}_{t-1}) ds_{t-1}. \quad (19)$$

This substitution yields rapid implementation if additionally the integral in (19) 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 m is also normal. In cases for which there is no analytical solution, a standard MC approximation can be used

$$\hat{f}_S(s_t|Y_{t-1}) \simeq \frac{1}{S} \sum_{i=1}^S f(s_t|s_{t-1}^i, Y_{t-1}). \quad (20)$$

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 (5) requires the construction of a continuous approximation to $f(s_{t-1}|Y_{t-1})$. Two obvious candidates are as follows. The first is the period- $(t-1)$ EIS sampler $m(s_{t-1}; \hat{a}_{t-1})$, under the implicit assumption that the corresponding weights $\omega(s_{t-1}; \hat{a}_{t-1})$ are near-constant, at least over the range of integration. The second is the use of a more flexible sampler, such as a mixture of Gaussian densities.

4.3.3 Degenerate transitions

When state transition equations include identities, corresponding transition densities are degenerate (or Dirac) in some of their components; this requires an adjustment to EIS implementation. Let s_t partition into $s_t = (p_t, q_t)$, such that

there is 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:

$$q_t \equiv \phi(p_t, p_{t-1}, q_{t-1}) = \phi(p_t, s_{t-1}). \quad (21)$$

The evaluation of $f(s_t|Y_{t-1})$ in (5) 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 (21). Note in particular that the presence of identities raises a conditioning issue known as the Borel-Kolmogorov paradox (see DeGroot, 2004, Section 3.10). This issue is resolved here by reinterpreting (21) 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

$$q_{t-1} = \psi(p_t, q_t, p_{t-1}) = \psi(s_t, p_{t-1}) \quad (22)$$

taking the limit of the integral in (5) as ε tends to zero, produces

$$f(s_t|Y_{t-1}) = \int J(s_t, p_{t-1}) f(p_t|s_{t-1}, Y_{t-1}) f(p_{t-1}, q_{t-1}|Y_{t-1})|_{q_{t-1}=\psi(s_t, p_{t-1})} dp_{t-1}, \quad (23)$$

where with $\|\cdot\|$ denoting the absolute value of a determinant,

$$J(s_t, p_{t-1}) = \left\| \frac{\partial}{\partial q_t} \psi(s_t, p_{t-1}) \right\|. \quad (24)$$

Note that (23) 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. Instead, (23) can be approximated by replacing $f(s_{t-1}|Y_{t-1})$ by $\bar{\omega}(\hat{a}_{t-1}) m(s_{t-1}|\hat{a}_{t-1})$:

$$\hat{f}(s_t|Y_{t-1}) = \int J(s_t, p_{t-1}) f(p_t|q_{t-1}, Y_{t-1}) m(p_{t-1}, q_{t-1}|\hat{a}_{t-1})|_{q_{t-1}=\psi(s_t, p_{t-1})} dp_{t-1}. \quad (25)$$

In this case, since $m(p_{t-1}, \psi(s_t, p_{t-1})|\hat{a}_{t-1})$ is not a sampler for $p_{t-1}|s_t$, (25) must be evaluated either by quadrature or its own EIS sampler.

4.4 Application to DSGE Models

As noted, the work of Fernandez-Villaverde and Rubio-Ramirez (2005, 2007) revealed that approximation errors associated with linear representations of DSGE models can impart significant errors in corresponding likelihood representations. As a remedy, they demonstrated the use of the particle filter for achieving likelihood evaluation for non-linear model representations. Here the EIS filter implementation will be explained using two workhorse models. The first is the standard two-state real business cycle (RBC) model; the second is a small-open-economy (SOE) model patterned after those considered, e.g., by Mendoza (1991) and Schmitt-Grohe and Uribe (2003), but extended to include six state variables.

Two data sets for each models are analyzed: an artificial data set generated from a known model parameterization; and a corresponding real data set. Thus in total four applications are considered, each of which poses a significant challenge to the successful implementation of a numerical filtering algorithm.

4.4.1 Example 1: Two-State RBC Model

The first application is to the simple DSGE model used by Fernandez-Villaverde and Rubio-Ramirez (2005) to demonstrate implementation of the particle filter. The model consists of a representative household that seeks to maximize the expected discounted stream of utility derived from consumption c and leisure l :

$$\max_{c_t, l_t} U = E_0 \sum_{t=0}^{\infty} \beta^t \frac{(c_t^\varphi l_t^{1-\varphi})^{1-\phi}}{1-\phi},$$

where (β, ϕ, φ) represent the household's subjective discount factor, degree of relative risk aversion, and the relative importance assigned to c_t and l_t in determining period- t utility.

The household divides its available time per period (normalized to unity) between labor n_t and leisure. Labor combines with physical capital k_t and a stochastic productivity term z_t to produce a single good ζ_t , which may be consumed or invested (ζ is used in place of the usual representation for output y – to avoid confusion with the use of y as representing the observable variables of

4.4 Application to DSGE Models

a generic state-space model). Investment i_t combines with undepreciated capital to yield k_{t+1} , thus the opportunity cost of period- t consumption is period- $(t + 1)$ capital. Collectively, the constraints faced by the household are given by

$$\begin{aligned}\zeta_t &= z_t k_t^\alpha n_t^{1-\alpha}, \\ 1 &= n_t + l_t, \\ \zeta_t &= c_t + i_t, \\ k_{t+1} &= i_t + (1 - \delta)k_t, \\ z_t &= z_0 e^{gt} e^{\omega_t}, \quad \omega_t = \rho\omega_{t-1} + \varepsilon_t,\end{aligned}$$

where $(\alpha, \delta, g, \rho)$ represent capital's share of output, the depreciation rate of capital, the growth rate of total factor productivity (TFP), and the persistence of innovations to TFP.

Optimal household behavior is represented as policy functions for $(\zeta_t, c_t, n_t, l_t, i_t)$ in terms of the state (k_t, z_t) . Given the policy function $i(k_t, z_t)$, the state-transitions equations reduce to

$$\begin{aligned}\left(1 + \frac{g}{1 - \alpha}\right) k_t &= i(k_{t-1}, z_{t-1}) + (1 - \delta)k_{t-1} & (26) \\ \ln z_t &= (1 - \rho) \ln(z_0) + \rho \ln z_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), & (27)\end{aligned}$$

and the observation equations are

$$x_t = x(k_t, z_t) + u_{x,t}, \quad x = \zeta, i, n, \quad u_{x,t} \sim N(0, \sigma_x^2). \quad (28)$$

Policy functions are expressed as Chebyshev polynomials in the state variables (k_t, z_t) , constructed using the projection method described in the Appendix to this Chapter, and in DeJong and Dave (2007).

Given the form of (27), it will be useful to represent state variables as logged deviations from steady state: $s_t = [\ln(k_t/k^*) \quad \ln(z_t/z^*)]'$. For ease of notation, hereafter $\ln(k_t/k^*)$ will denote k_t , and $\ln(z_t/z^*)$ will denote z_t . In addition, given the form of (28), y_t is defined as $y_t = [\zeta_t \quad i_t \quad n_t]'$. All subsequent formulas should be read in accordance with these representations.

To obtain the predictive density associated with (26) and (27), note that since (26) is an identity, the transition density of the system is degenerate in k_t . Thus

4. NONLINEAR STATE-SPACE MODELS

(26) is inverted yielding

$$k_{t-1} = \psi(k_t, z_{t-1}), \quad (29)$$

$$J(k_t, z_{t-1}) = \left\| \frac{\partial}{\partial k_t} \psi(k_t, z_{t-1}) \right\|, \quad (30)$$

and express the predictive density as

$$f(s_t | Y_{t-1}) = \int J(k_t, z_{t-1}) f(z_t | s_{t-1}) f(s_{t-1} | Y_{t-1}) |_{k_{t-1}=\psi(k_t, z_{t-1})} dz_{t-1}. \quad (31)$$

From (27), note that

$$f(z_t | s_{t-1}) \sim N_1 \left(\begin{bmatrix} 0 & \rho \end{bmatrix} s_{t-1}, \sigma_\varepsilon^2 \right),$$

with $N_r(\cdot)$ denoting an r -dimensional normal distribution. Finally, as the inversion of Chebyshev polynomials is awkward, (29) and (30) are approximated using third-order polynomials in (k_t, z_{t-1}) .

With the predictive density established, the time- t likelihood is standard:

$$f(y_t | Y_{t-1}) = \int f(y_t | s_t, Y_{t-1}) f(s_t | Y_{t-1}) ds_t, \quad (32)$$

where from (28),

$$f(y_t | s_t, Y_{t-1}) \sim N_3(\mu(s_t), V), \quad (33)$$

$$\mu(s_t) = \begin{bmatrix} \zeta(s_t) \\ i(s_t) \\ n(s_t) \end{bmatrix}, \quad V = \begin{bmatrix} \sigma_y^2 & 0 & 0 \\ 0 & \sigma_i^2 & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}.$$

To achieve likelihood evaluation in period t , the approach is to construct a normally distributed EIS sampler $m(s_t; \widehat{a}_t)$ for the integrand $f(y_t | s_t, Y_{t-1}) \cdot f(s_t | Y_{t-1})$ in (32). In so doing, $f(s_t | Y_{t-1})$ is represented using the constant-weight approach to approximation described above. That is, the time- $(t-1)$ sampler $m(s_{t-1}; \widehat{a}_{t-1})$ is used as a stand-in for $f(s_{t-1} | Y_{t-1})$, yielding

$$f(s_t | Y_{t-1}) \simeq \int J(k_t, z_{t-1}) f(z_t | s_{t-1}) m(s_{t-1}; \widehat{a}_{t-1}) |_{k_{t-1}=\psi(k_t, z_{t-1})} dz_{t-1}. \quad (34)$$

The process is initialized by constructing $f(s_0)$ as the unconditional distribution of the Kalman filter associated with a linear approximation of the model, which is only used to construct the EIS sampler for the first time period $m(s_1; \widehat{a}_1)$.

4.4 Application to DSGE Models

It proceeds via forward recursion, taking $m(s_{t-1}; \widehat{a}_{t-1})$ as an input, and passing $m(s_t; \widehat{a}_t)$ to the subsequent period. Full details follow.

Consider first the evaluation of (34). Representing $m(s_t; \widehat{a}_t)$ as

$$m(s_t; \widehat{a}_t) \sim N_2(\mu, \Omega),$$

and with $f(z_t|s_{t-1})$ distributed as

$$f(z_t|s_{t-1}) \sim N_1\left(\begin{bmatrix} 0 & \rho \end{bmatrix} s_{t-1}, \sigma_\varepsilon^2\right),$$

$f(z_t|s_{t-1}) m(s_{t-1}; \widehat{a}_{t-1})$ combines to form the joint density

$$\begin{pmatrix} z_t \\ s_{t-1} \end{pmatrix} \sim N_3\left(\begin{pmatrix} 0 & \rho \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \mu, V\right), \quad (35)$$

$$V = \sigma_\varepsilon^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & \rho \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \Omega \begin{pmatrix} 0 & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix}.$$

As (34) must be evaluated for each candidate s_t used to calculate (32), one must transform (35) into a distribution over $(s_t, z_{t-1})'$. Approximating (29) linearly as $k_{t-1} = a_k k_t + b_k z_{t-1}$, which implies

$$\begin{pmatrix} s_t \\ z_{t-1} \end{pmatrix} = A \begin{pmatrix} z_t \\ s_{t-1} \end{pmatrix},$$

$$A = \begin{pmatrix} 0 & 1/a_k & -b_k/a_k \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad |A| = 1/|a_k|,$$

it is possible to express $(z_t \ s_{t-1})'$ as a function of $(s_t \ z_{t-1})'$. This expression, coupled with (35), yields the joint density

$$\begin{aligned} f^*(s_t, z_{t-1}) &\sim N_3(m, \Sigma), & \Sigma &= AVA', & (36) \\ m &= B\mu, \\ B &= A \begin{bmatrix} 0 & \rho \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

4. NONLINEAR STATE-SPACE MODELS

Finally, partitioning (36) into a product of two densities, one for s_t and one for $z_{t-1}|s_t$:

$$f_1^*(s_t) \sim N_2(m_1, \Sigma_{11}), \quad f_2^*(z_{t-1}|s_t) \sim N_1(m_{2.1} + \Delta_{21}s_t, \sigma_{22.1}),$$

where

$$m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \sigma_{22} \end{pmatrix},$$

with Σ_{11} 2×2 , Σ_{12} 2×1 , Σ_{21} 1×2 , and σ_{22} 1×1 , and

$$\Delta_{21} = \Sigma_{21} (\Sigma_{11})^{-1}, \quad m_{2.1} = m_2 - \Delta_{21}m_1, \quad \sigma_{22.1} = \sigma_{22} - \Sigma_{21} (\Sigma_{11})^{-1} \Sigma_{12}.$$

Having accomplished these steps, (34) is approximately

$$f(s_t|Y_{t-1}) \approx f_1^*(s_t) \int \frac{J(s_t, z_{t-1})}{|a_k|} f_2^*(z_{t-1}|s_t) dz_{t-1}, \quad (37)$$

where since

$$|\Sigma|^{-1/2} = \frac{1}{|a_k|} |V|^{-1/2},$$

the term $|a_k|$ enters via the usual change-of-variables formula. For each candidate s_t that enters into the approximation of (32), $f_2^*(z_{t-1}|s_t)$ is used as an EIS sampler, and approximate (37) as

$$f(s_t|Y_{t-1}) \approx \left(\frac{f_1^*(s_t)}{|a_k|} \right) \frac{1}{S} \sum_{i=1}^S J(s_t, z_{t-1}^i), \quad (38)$$

where $\{z_{t-1}^i\}_{i=1}^S$ are simulated drawings from $f_2^*(z_{t-1}|s_t)$.

Turning to the approximation of (32), this is straightforward once a reliable initial EIS sampler is constructed. Such an initial sampler should be constructed as a close approximation of the integrand $f(y_t|s_t, Y_{t-1}) \cdot f(s_t|Y_{t-1})$. Towards this end, $f_1^*(s_t)$ is used in place of $f(s_t|Y_{t-1})$, to construct a linear Gaussian approximation of $f(y_t|s_t, Y_{t-1})$. Recall from (33) that $f(y_t|s_t, Y_{t-1})$ is already Gaussian, but with a non-linear mean function, say $\mu(s_t)$. Approximating this function as $\mu(s_t) \approx r + Ps_t$, yields

$$f^*(y_t|s_t, Y_{t-1}) \sim N_3(r + Ps_t, V).$$

4.4 Application to DSGE Models

Combining these approximations, and letting $Q = V^{-1}$, $H = \Sigma_{11}^{-1}$, yields the initial sampler

$$\begin{aligned} m(s_t; a_t^0) &= f_1^*(s_t) f^*(y_t | s_t, Y_{t-1}) \\ &\sim N(\mu^0, \Sigma^0), \\ \mu^0 &= (H + P'QP)^{-1} [Hm_1 + P'Q(y_t - r)], \\ \Sigma^0 &= (H + P'QP)^{-1}. \end{aligned} \tag{39}$$

To summarize, EIS implementation is achieved for the two-state RBC model as follows.

Model Representation

- Policy functions $x(k_t, z_t)$, $x = (\zeta, c, i, n, l)$, expressed as Chebyshev polynomials in $s_t = [k_t \ z_t]'$, are constructed via the projection method.
- With the law of motion for capital given by $k_t = i(k_{t-1}, z_{t-1}) + (1 - \delta)k_{t-1}$, solve for k_{t-1} to obtain

$$k_{t-1} = \psi(k_t, z_{t-1}), \quad J(k_t, z_{t-1}) = \frac{\partial}{\partial k_t} \psi(k_t, z_{t-1}),$$

represented as third-order polynomials in (k_t, z_{t-1}) . The linear approximation $k_{t-1} = a_k k_t + b_k z_{t-1}$ is also constructed.

Likelihood Evaluation

- The EIS sampler $m(s_{t-1}; \widehat{a_{t-1}})$ serves as an input in the construction of the period- t likelihood function. In period 1, $m(s_0; \widehat{a_0}) \equiv \widehat{f}(s_0)$ is constructed as the unconditional distribution of the Kalman filter associated with a linear approximation of the model.
- To approximate the integrand $f(y_t | s_t, Y_{t-1}) \cdot f(s_t | Y_{t-1})$ in (32), an initial sampler $m(s_t; a_t^0)$ is constructed as in (39):

$$m(s_t; a_t^0) = f_1^*(s_t) f^*(y_t | s_t, Y_{t-1})$$

- Using drawings $\{s_{t,0}^i\}_{i=1}^R$ obtained from $m(s_t; a_t^0)$, \widehat{a}_t is obtained as the solution to (14). This entails the computation of

$$\varphi_t(s_{t,l}^i) = f(y_t | s_{t,l}^i, Y_{t-1}) f(s_{t,l}^i | Y_{t-1}),$$

where $f(y_t | s_t, Y_{t-1})$ is given in (33), and $f(s_t | Y_{t-1})$ is approximated as indicated in (38).

4. NONLINEAR STATE-SPACE MODELS

Table 4.1: Parameter Values, RBC Model

	α	β	ϕ	φ	δ	ρ	σ_ε	σ_y	σ_i	σ_n
Artificial	0.4	0.99	2	0.357	0.02	0.95	0.007	$1.58e-4$	$8.66e-4$	0.0011
Actual	0.324	0.997	1.717	0.390	0.006	0.978	0.020	0.045	0.038	0.015

- Having constructed $m(s_t; \hat{a}_t)$, $\hat{f}(y_t|Y_{t-1})$ is approximated as indicated in (15).
- The sampler $m(s_t; \hat{a}_t)$ is passed to the period- $(t+1)$ step of the algorithm. The algorithm concludes with the completion of the period- T step.

To demonstrate the performance of the EIS filter in this setting, Monte Carlo experiments using two data sets were conducted. The first is an artificial data set consisting of 100 realizations of $\{\zeta_t, i_t, n_t\}$ generated from the RBC model. This was constructed by Fernandez-Villaverde and Rubio-Ramirez (2005) under the model parameterization presented in the first row of Table 4.1. The second consists of actual quarterly observations on $\{\zeta_t, i_t, n_t\}$ used by Fernandez-Villaverde and Rubio-Ramirez to estimate the RBC model using the particle filter. The data are quarterly, span 1964:I-2003:II (158 observations), and were detrended using the Hodrick-Prescott filter. Posterior means of the estimates they obtained using this data set are presented in the second row of Table 4.1. Both data sets are available for downloading at <http://qed.econ.queensu.ca/jae/2005-v20.7/fernandez-rubio/>

Each data set poses a distinct challenge to efficient filtering. In the artificial data set, note from Table 4.1 that the standard deviations of the measurement errors ($\sigma_y, \sigma_i, \sigma_n$) are small relative to σ_ε , which as noted above can lead to problems associated with sample impoverishment. In the real data set, the investment series contains two outliers: the values at 1976:III and 1984:IV, which lie 7.7 and 4.7 standard deviations above the sample mean. Outliers can induce bias in likelihood estimates associated with the particle filter. Both of these challenges are overcome via implementation of the EIS filter, as it will be now demonstrated.

Using both data sets, a Monte Carlo experiment was conducted under which 1,000 approximations of the likelihood function were produced (evaluated at Table 4.1 parameter values) for both the particle and EIS filters using 1,000 different sets of random numbers. Differences in likelihood approximations across sets of random numbers are due to numerical approximation errors. Following

4.4 Application to DSGE Models

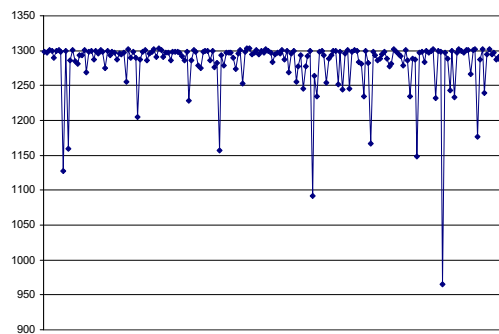
Fernandez-Villaverde and Rubio-Ramirez, the particle filter was implemented using $N = 60,000$, requiring 40.6 seconds of CPU time per likelihood evaluation on a 1.8 GHz desktop computer using GAUSS for the artificial data set, and 63 seconds for the real data set. The EIS filter was implemented using $N = R = 20$, $S = 10$, with one iteration used to construct \hat{a}_t ; this required 0.22 seconds per likelihood evaluation for the artificial data set, and 0.328 seconds for the real data set.

Considering first the artificial data set, the mean and standard deviation of the 1,000 log-likelihood approximations obtained using the particle filter are (1, 285.51; 33.48), and (1, 299.81; 0.00177) using the EIS filter (the likelihood value obtained using the Kalman filter and the log-linear model is 1,300.045). Thus, compared to the particle filter, the EIS filter reduces numerical approximation errors by four orders of magnitude in this application. Figure ?? plots the first 200 likelihood approximations obtained using both filters and a rescaled plot of the EIS approximations in order to enhance visibility. Note that the particle-filter approximations (Figure 4.3(a)) often fall far below the EIS sample mean of 1,299.81 (by more than 50 on the log scale in twenty instances, and by more than 100 in eight instances); this largely accounts for the distinct difference in sample means obtained across methods. But as the figure indicates, even abstracting from the occasional large likelihood crashes suffered by the particle filter, the EIS filter is extremely precise: the maximum difference in log-likelihood values it generates is less than 0.012 (Figure 4.3(c)), while differences of 10 are routinely observed for the particle filter.

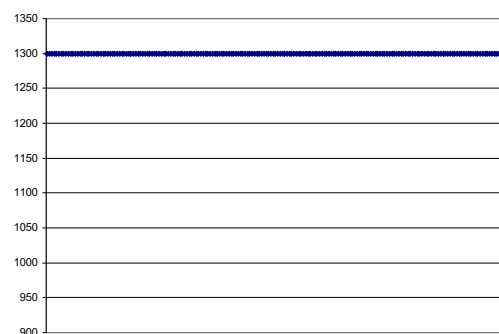
Hereafter, the differences observed between sample means of log-likelihood values obtained using the particle and EIS filters will be referred to as reflecting bias associated with the particle filter. This presumes that the values associated with the EIS filter closely represent “truth”. This presumption is justified in a number of ways, in this experiment and each of those that follow. First, the small numerical approximation errors associated with the EIS filter indicate virtual replication across sets of random numbers. Second, as the values of (N, R) used to implement the EIS filter are increased, resulting mean log-likelihood approximations remain virtually unchanged, while numerical errors are inversely proportional to $N^{-1/2}$, as expected. Finally, when the EIS filter is implemented using linear model approximations, the log-likelihood values obtained match those

4. NONLINEAR STATE-SPACE MODELS

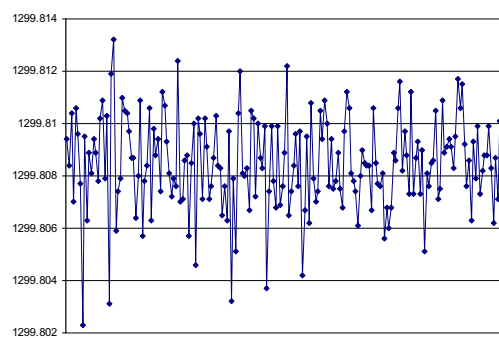
Figure 4.2: Log-Likelihood Approximations



(a) Particle Filter



(b) EIS Filter



(c) EIS Filter Rescaled

4.4 Application to DSGE Models

produced by the Kalman filter almost exactly (virtually to the limits of numerical precision).

Figure ?? provides an illustration and diagnosis of the problems faced by the particle filter in this application. The focus here is on a representative Monte Carlo replication generated as described above. Each one of the figures represents a distinct scenario observed routinely across time periods within this replication. Figure 4.4(a) corresponds with $t = 53$, and Figure 4.4(b) with $t = 18$.

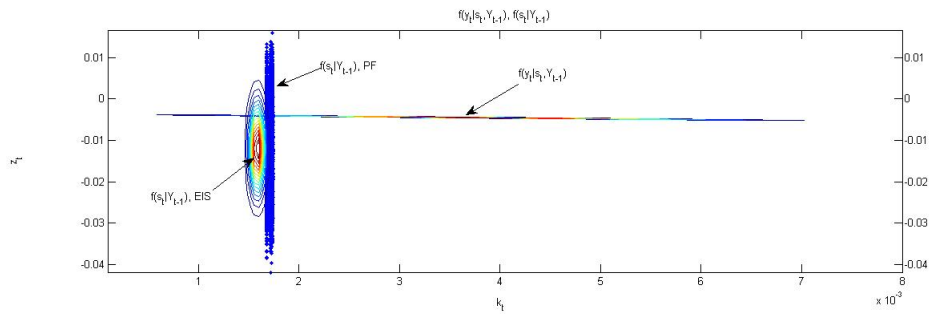
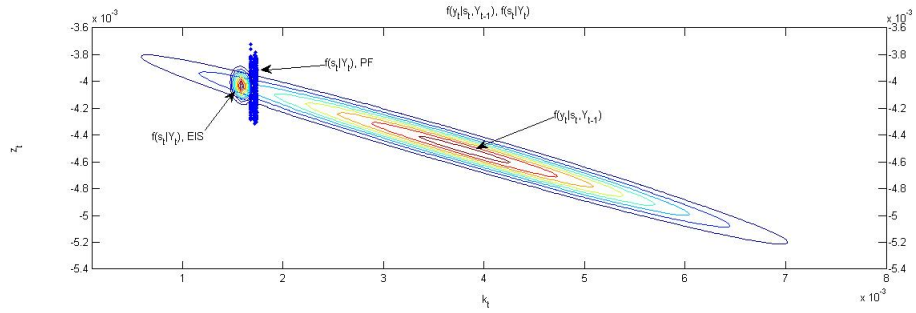
Each Figure contains two graphs, both of which depict z_t on the vertical axis and k_t on the horizontal axis. The measurement density $f(y_t|s_t, Y_{t-1})$ is the large thin ellipse depicted in both graphs (differences in vertical scales across graphs account for differences in its appearance). In the bottom graph, the swarm of dots comprises the particle-filter representation of $f(s_t|Y_{t-1})$, and the wide ellipse comprises the EIS representation of $f(s_t|Y_{t-1})$. In the upper graph, the swarm of dots comprises the particle-filter representation of $f(s_t|Y_t)$; particles in the upper swarm were obtained by sampling repeatedly from the bottom swarm, with probabilities assigned by the measurement density. The upper graph also depicts the EIS representation of $f(s_t|Y_t)$ (small ellipse).

Beginning with period 53, note that the vast majority of particles in the bottom graph are assigned negligible weight by the measurement density, and are thus discarded in the resampling step. Specifically, only 407 particles, or 0.68% of the total candidates, were re-sampled at least once in this instance. The average (across time periods) number of re-sampled particles is 350, or 0.58% of the total. This phenomenon reflects the sample impoverishment problem noted above. It results from the ‘blindness’ of proposals generated under the particle filter algorithm, and accounts for its numerical inaccuracy.

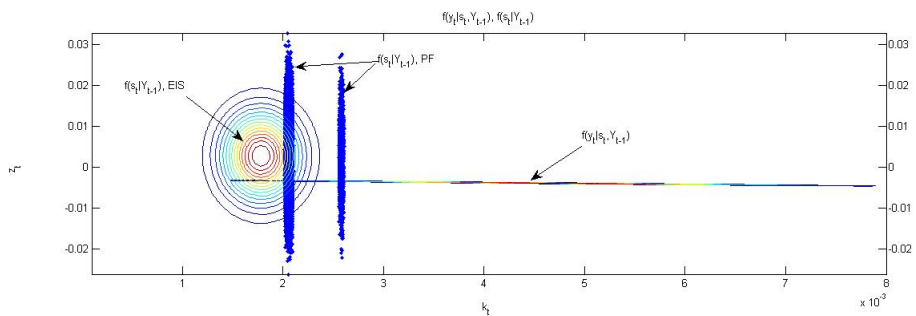
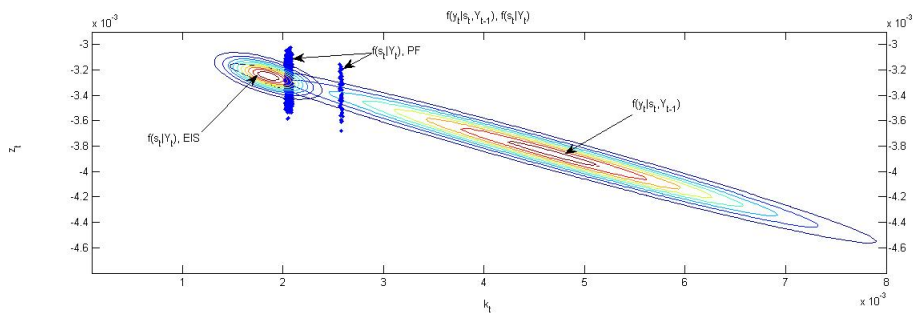
As noted, the small ellipse depicted in the upper graph is the EIS representation of $f(s_t|Y_t)$. The difference between this and the corresponding particle-filter representation reflects a second problem suffered by the particle filter in this application: there is non-trivial bias in the filtered values of the state it produces. This also reflects the ‘blindness’ problem, coupled with the fact that alternative proposals for s_t cannot be re-generated in light of information embodied in y_t . (Note from the figure that this bias is not easily eliminated through an increase in the number of particles included in the proposal swarm, since the probability that the proposal density will generate particles centered on the EIS representation

4. NONLINEAR STATE-SPACE MODELS

Figure 4.3: Sample Impoverishment in Practice I



(a) Time Period 53



(b) Time Period 18

4.4 Application to DSGE Models

of $f(s_t|Y_t)$ is clearly miniscule.) As described above, under suitable initialization the EIS filter avoids these issues by generating proposals from an importance density tailored as the optimal global approximation of the targeted integrand $f(y_t|s_t, Y_{t-1}) \cdot f(s_t|Y_{t-1})$.

Regarding period 18, note that the representations of both $f(s_t|Y_{t-1})$ and $f(s_t|Y_t)$ generated using the particle filter are discontinuous in k_t , a spurious phenomenon that occurs frequently through the sample. This exacerbates the bias associated with filtered values of the state, and contributes to a final problem associated with the particle filter illustrated in Figure 4.4.

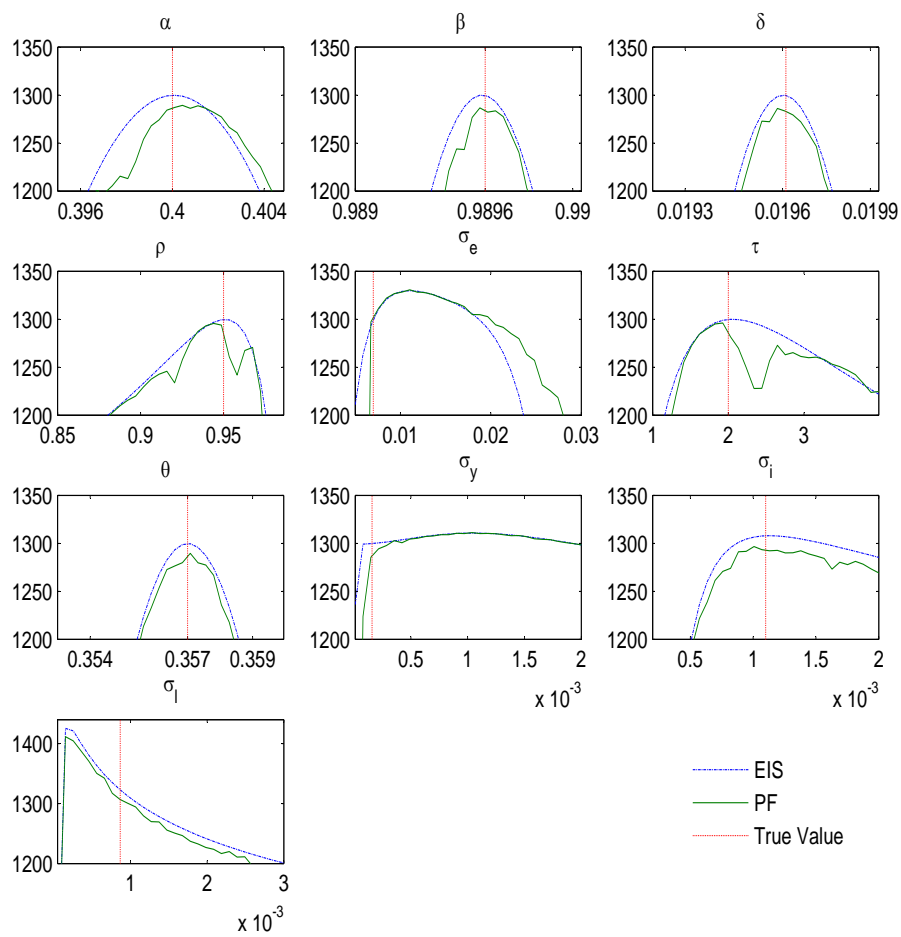
Like its predecessor, Figure 4.4 was produced using a representative Monte Carlo replication. It depicts an approximation of the log-likelihood surface over a given parameter obtained by holding the remaining fixed at their true values, and varying the chosen one above and below the true value. Two surfaces are depicted in each graph: those associated with the particle, and EIS filters. The particle and EIS surfaces were produced with common random numbers, so that changes in parameters serve as the lone source of variation in log-likelihoods. This means that the kinks and local maxima that can be observed in the particle filter surfaces are spurious.

Note that for α while the surface associated with the EIS filter is continuous and peak at the true value of 0.4, the surface associated with the particle filter is discontinuous and has a slightly rightshifted peak. Thus in addition to being numerically inefficient and producing biased filtered values of the state, the particle filter generates likelihood surfaces that are spuriously discontinuous in the underlying parameters of the model, rendering as problematic the attainment of likelihood-based model estimates.

To illustrate the difficulties created by the discontinuities in the likelihood approximations of the particle filter and how the EIS filter is able to overcome them, maximum likelihood estimation of the RBC model was performed using both filters based on the derivative free Nelder-Meade Simplex algorithm (see Lagarias et al., 1998). The estimates are reported in Table 4.2. The EIS Filter estimates are based on $N = R = 20$ draws for the likelihood integral approximation given by (32), and $S = 10$ draws for the integral approximation to the predictive density in (38). The estimates using the particle filter are based on a swarm with 60,000 particles. The asymptotic standard errors for the EIS filter

4. NONLINEAR STATE-SPACE MODELS

Figure 4.4: Likelihood Cuts



4.4 Application to DSGE Models

estimates were easily calculated using central differences. However, given the discontinuities of the particle filter's likelihood approximation, it was not possible to compute its Hessian matrix even when the size of the swarm was increased to 200,000 particles, making clear that the discontinuities are not easily resolved via the increase of the MC sample size.

Table 4.2: Estimates of RBC Model using simulated data.

		EIS Filter		Particle Filter	
	True Values	ML Est	STD	ML Est	STD
α	0.4000	0.4000	$4.73e-05$	0.3993	-
β	0.9896	0.9894	$5.99e-05$	0.9899	-
δ	0.0196	0.0197	$1.47e-05$	0.0193	-
ρ	0.9500	0.9551	$8.90e-04$	0.9460	-
σ_ε	0.0070	0.0125	$1.97e-04$	0.0110	-
τ	2.0000	1.9898	$2.91e-03$	2.3029	-
θ	0.3570	0.3569	$1.28e-05$	0.3567	-
σ_y	$1.58e-04$	$1.40e-04$	$7.65e-06$	$7.98e-05$	-
σ_i	0.0011	0.001136	$1.20e-05$	0.001120	-
σ_l	$8.66e-04$	$1.84e-04$	$6.49e-06$	$1.82e-04$	-
$\log lik$		1470.41230		1468.642181	

Turning to the experiment conducted using the actual data set, note that in this case there are non-trivial differences between the log-likelihood values associated with the Kalman and EIS filters. The mean and standard deviation of the 1,000 log-likelihood approximations obtained using the EIS filter are $(921.56, 0.056)$, while the log-likelihood value associated with the Kalman filter is 928.06. The explanation for this difference is as follows. Since there are no outliers in the artificial data set, deviations from steady state are relatively small, thus the linear model approximations employed in implementing the Kalman filter are relatively accurate. However, accuracy breaks down when deviations from steady state are large (i.e., in the presence of outliers). Indeed, when the EIS filter is implemented using linear model approximations, differences in likelihoods produced by the EIS and Kalman filters virtually disappear (becoming at most $1.58e-9$ in 1976:IV).

4. NONLINEAR STATE-SPACE MODELS

Second, compared with the first example, differences between the particle and EIS filters are relatively modest in this case. The mean and standard deviation of the 1,000 log-likelihood approximations obtained using the particle filter are $(926.84, 0.1958)$. Thus the difference in sample means is $5\frac{1}{4}$ in this case, compared with more than 14 in the first example, while the difference in standard deviations is by a factor of 3.5, as opposed to four orders of magnitude in the first example. The explanation for this is that measurement densities are relatively diffuse in this case, thus the sample impoverishment problem is less acute in general.

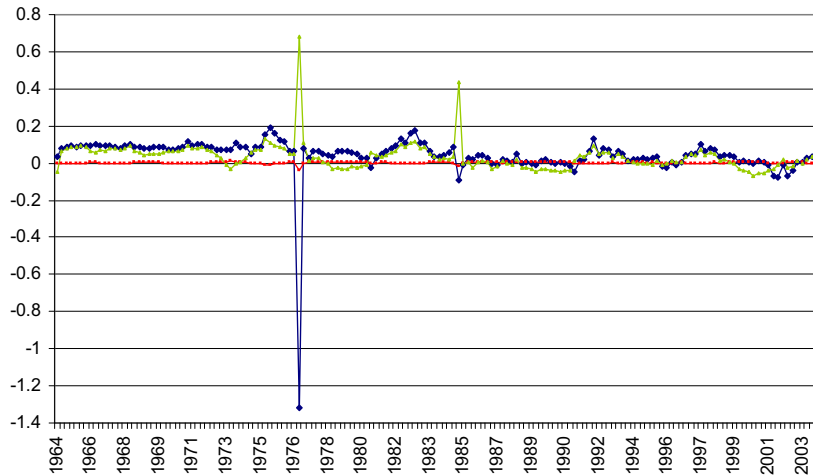
As the primary interest in this example is on the impact of outliers, Figures 4.6(a) and 4.6(b) illustrate the relative accuracy and precision of the Kalman, particle and EIS filters on a date-by-date basis. Figure 4.6(a) illustrates average (across CRNs) deviations from pseudo-true values of log-likelihood values associated with all three filters; Figure 4.6(b) illustrates MC standard deviations (across CRNs) of log-likelihood approximations associated with the EIS and particle filters. Thus Figure 4.6(a) provides a characterization of bias, and Figure 4.6(b) numerical precision. To generate pseudo-true values, log-likelihood values using the EIS filter implemented by setting (N, R) as $(2000, 1000)$ are computed.

Absent the two outlier periods, the Kalman and particle filters exhibit non-trivial but relatively moderate bias. The average and total (across dates) absolute deviations from pseudo-true values of log-likelihoods obtained using the EIS filter are $(0.002, 0.358)$, compared with $(0.057, 8.955)$ using the Kalman filter and $(0.043, 6.721)$ using the EIS filter. However, bias is distinctly more pronounced in the presence of the outliers, particularly for that observed in 1976:III. In this period absolute deviations in log-likelihoods are 0.041, 1.321 and 0.682 for the EIS, Kalman, and particle filters.

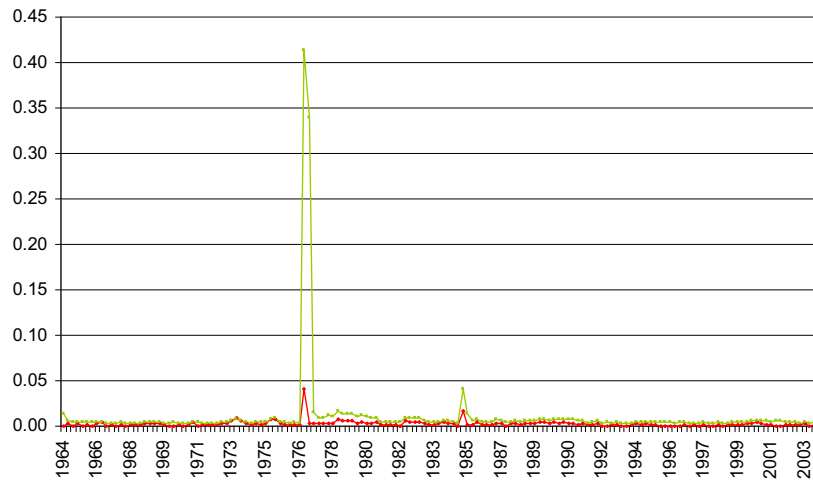
The outliers also have a distinct impact on the numerical precision of the particle filter, as the Figure 4.6(b) illustrates. In particular, while the average (across dates) MC standard deviation calculated for the particle filter is 0.0077 absent the two outliers, the standard deviation jumps to 0.41 in 1976:III, and 0.04 in 1984:IV. Associated values for the EIS filter are 0.0041, 0.03, and 0.01.

An illustration of the source of these biases and numerical inaccuracies is provided in Figures 4.6 and 4.7. The former focusses on the Kalman filter, the latter on the EIS filter; both pertain to the period 1976:III, and were generated using a single set of CRNs.

Figure 4.5: Date-by-Date Filter Comparisons



(a) Deviations of log-Likelihood from Pseudo-True Values



(b) MC Standard Deviations of Log-Likelihoods

*Date-by-Date Filter Comparisons, RBC Model, Actual Data
 Blue: Kalman Filter; Green: Particle Filter; Red: EIS Filter*

4. NONLINEAR STATE-SPACE MODELS

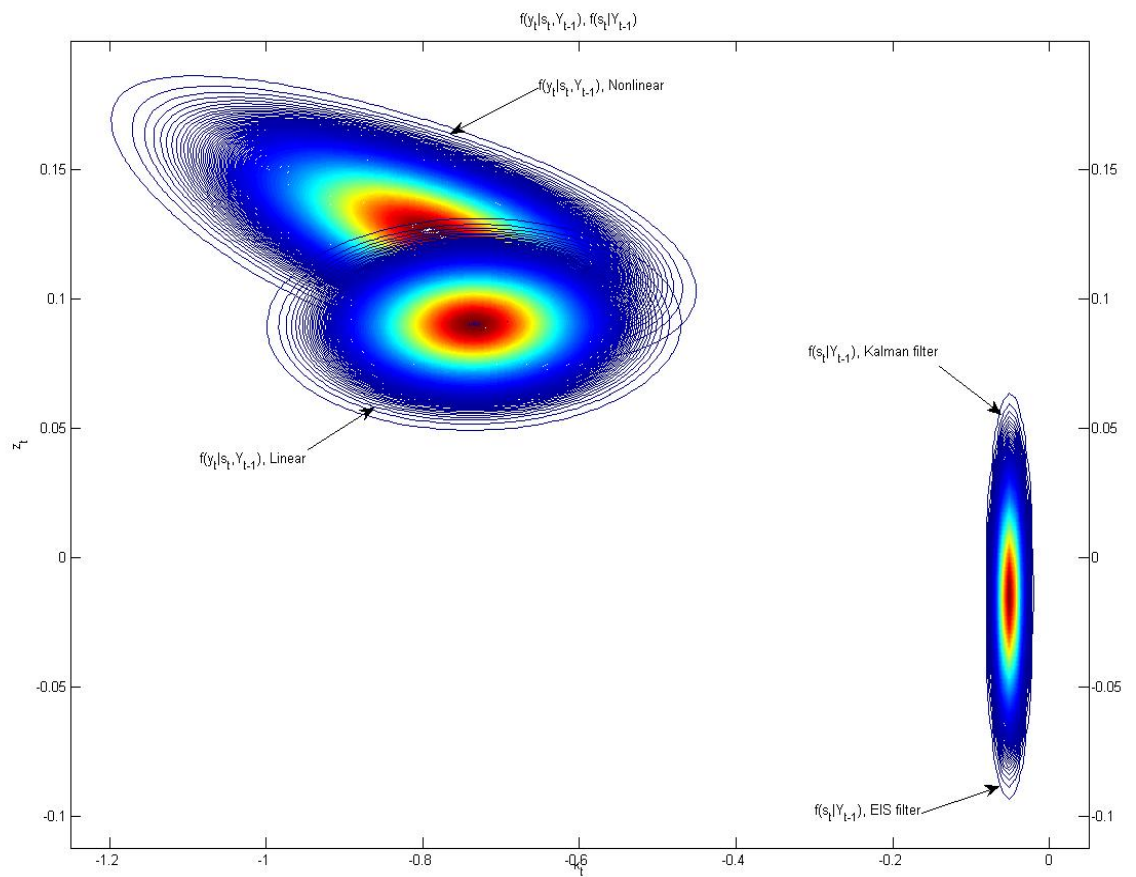
Figure 4.6 illustrates the measurement density $f(y_t|s_t, Y_{t-1})$ and predictive density $f(s_t|Y_{t-1})$ associated with both the Kalman and EIS filters. Recall that the integral of these densities over the sample space yields the likelihood function $f(y_t|Y_{t-1})$. Note first that the predictive densities associated with the two filters are virtually indiscernible, and lie near the steady state location of the diagram (indicated by values of zero for both k_t and z_t). This reflects the fact that the time- $(t-1)$ observations y_{t-1} lie relatively close to their steady state values. In contrast, due to the realization of the time- t outlier, the measurement densities lie far from steady state, which accounts for the distinct difference in their shapes and locations: the quality of the linear model approximation associated with the Kalman filter deteriorates as the state variables deviate further from their steady state values, thus its characterization of $f(y_t|s_t, Y_{t-1})$ deviates from that of the EIS filter. Since the Kalman-filter representation of $f(y_t|s_t, Y_{t-1})$ is relatively tightly distributed, its height is much lower than that of the EIS filter at the location of $f(s_t|Y_{t-1})$, thus accounting for the negative bias associated with the Kalman filter.

Turning to the particle filter, the construction of Figure 4.7 mirrors that of Figures 4.4(a) and 4.4(b), with one small exception. In the top panel, the swarm comprising the particle-filter representation of $f(s_t|Y_t)$ has individual elements that vary by size in direct proportion to the number of times they were resampled from the swarm $f(s_t|Y_{t-1})$ in the bottom diagram. This is done to more clearly illustrate how the particle filter yields a biased approximation of $f(s_t|Y_t)$ in this case. The largest particle in the figure is more than 10,000 times larger than the smallest, thus it has 10,000 times more weight in representing $f(s_t|Y_t)$. The uneven size of the particles in the swarm illustrates the sample impoverishment that results from the outlier.

From the bottom panel of Figure 4.7, note that the particle-filter representation of $f(s_t|Y_{t-1})$ is left-shifted relative to the EIS-filter representation. Since the particle-filter representation has a discrete and fixed support, this left-shift persists in the re-sampling step under which the filtering density $f(s_t|Y_t)$ is obtained. The upshot is that the particle-filter representation of $f(s_t|Y_t)$ provides insufficient (virtually non-existent) coverage of the north-east portion of the upper diagram: its representation of $f(s_t|Y_t)$ is biased.

4.4 Application to DSGE Models

Figure 4.6: Linear vs. Nonlinear Measurement Densities



Kalman vs. EIS Filter, Actual Data, 1976:III

4. NONLINEAR STATE-SPACE MODELS

The left-shift in $f(s_t|Y_{t-1})$ also induces bias in the approximation of the likelihood function produced by the particle filter. To see why, recall that $f(s_t|Y_{t-1})$ serves as the importance-sampling distribution used by the particle filter: its approximation of the likelihood function $f(y_t|Y_{t-1})$ is given by the average value of the measurement density $f(y_t|s_t, Y_{t-1})$ evaluated at each particle in the swarm $f(s_t|Y_{t-1})$. Since the particle-filter representation of $f(s_t|Y_{t-1})$ is left-shifted, and lies spuriously close to $f(y_t|s_t, Y_{t-1})$, the resulting likelihood approximation it produces is biased upwards.

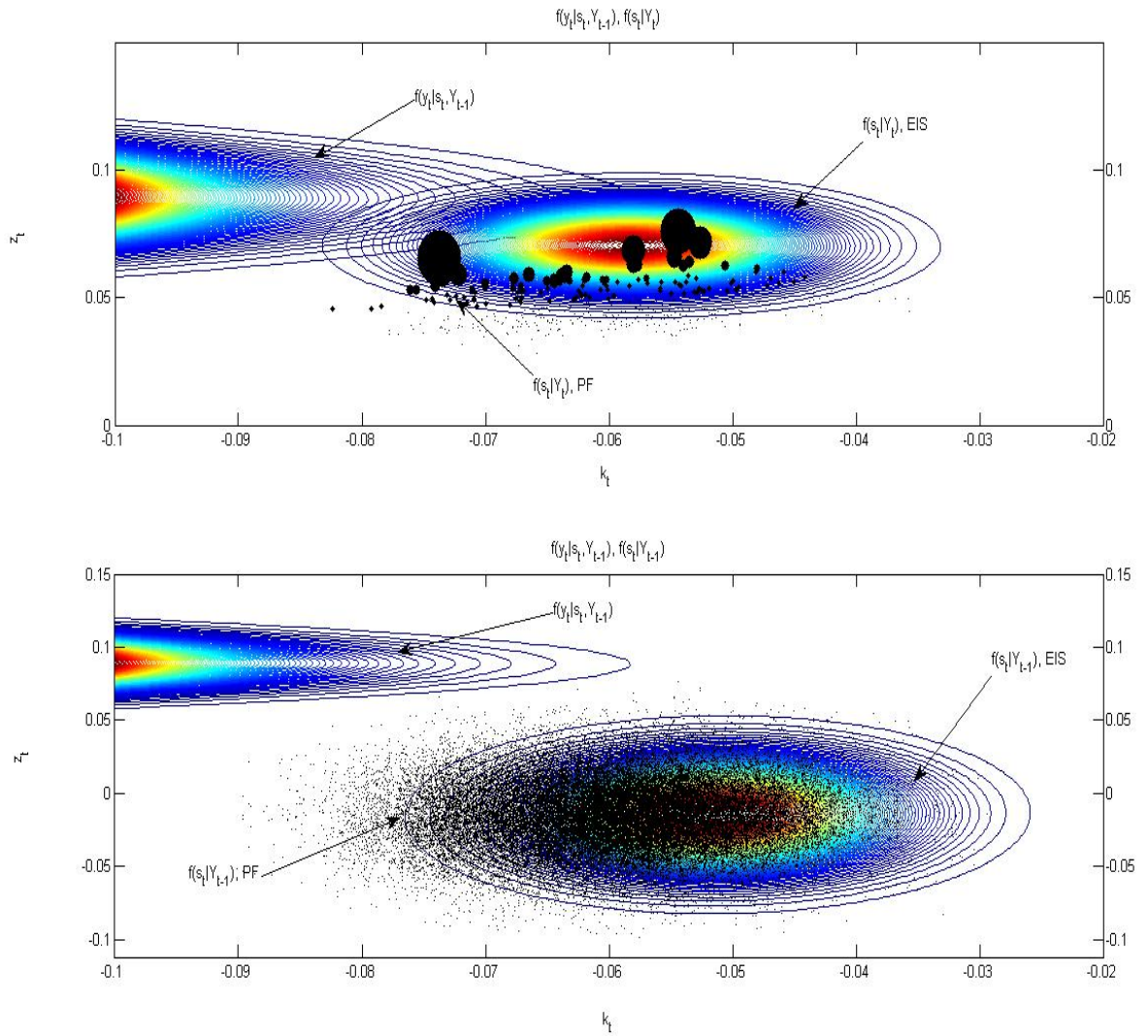
As a final note on Figure 4.7, the filtering density $f(s_t|Y_t)$ associated with the particle filter depicted in the upper diagram helps illustrate sources of numerical imprecision that plague the particle filter given the realization of an outlier. Only a handful of particles are assigned appreciable weight in this representation. Moreover, the exact location of these particles owes much to random chance: alternative sets of CRNs give rise to subtle locational shifts that are magnified due to the imbalanced weight assigned to a select few particles. Thus the spikes in numerical error associated with the outlier dates evident in Figures 4.6(a) and 4.6(b) are not surprising.

As emphasized by Richard and Zhang (2007), it is important to distinguish between the numerical error associated with a given approximation technique (quantified using the MC standard errors described above), and the sampling error associated with the statistic being approximated (in this case, the log-likelihood function). To characterize sampling error, two additional experiments were conducted. In both, a data generation process (DGP) using a parameterization of the RBC model were constructed, and 100 artificial data sets consisting of time-series observations of (ζ, i, n) of length T were generated. For each artificial data set, the EIS filter was implemented using $(N, R) = (200, 100)$ to obtain 100 approximations of the log-likelihood function. The standard deviation of the log-likelihoods calculated in this manner serves as an estimate of the statistical sampling error associated with this summary statistic.

The DGPs employed in the two experiments were tailored to the empirical applications described above. The first was constructed using the parameters reported in the second row of Table 4.1, with $T = 100$; the second using the parameters reported in the third row of Table 4.1, with $T = 158$. The first yielded an estimated sampling error of 16.48; the second 17.99. For comparison,

4.4 Application to DSGE Models

Figure 4.7: Sample Impoverishment in Practice II



Particle vs. EIS Filter, Actual Data, 1976:III

4. NONLINEAR STATE-SPACE MODELS

recall that the corresponding MC standard errors associated with the particle filter are 33.48 and 0.1958, while those associated with the EIS filter are 0.00177 and 0.0557. This comparison indicates that the particle filter is an unreliable tool for assessing statistical uncertainty in the context of the first example, since its associated numerical errors are first-order comparable to the associated statistical errors targeted for approximation.

To conclude, tightly-distributed measurement distributions and sample outliers are troublesome sources of numerical error and bias that can plague applications of the particle filter, but that can be overcome via application of the EIS filter. Now an application of the EIS filter in a second example model featuring an expanded state space will be demonstrated.

4.4.2 Example 2: Six-State Small Open Economy Model

This application is to a small-open-economy (SOE) model patterned after those considered, e.g., by Mendoza (1991) and Schmitt-Grohe and Uribe (2003). The model consists of a representative household that seeks to maximize

$$U = E_0 \sum_{t=0}^{\infty} \theta_t \frac{[c_t - \varphi_t \omega^{-1} n_t^\omega]^{1-\gamma} - 1}{1-\gamma}, \quad \omega > 0, \quad \gamma \geq 0,$$

where φ_t is a preference shock that affects the disutility generated by labor effort, introduced, e.g., following Smets and Wouters (2003). Following Uzawa (1968), the discount factor θ_t is endogenous and obeys

$$\begin{aligned} \theta_{t+1} &= \beta(\tilde{c}_t, \tilde{n}_t) \theta_t, & \theta_0 &= 1, \\ \beta(\tilde{c}_t, \tilde{n}_t) &= [1 + \tilde{c}_t - \omega^{-1} \tilde{n}_t^\omega]^{-\psi}, & \psi &> 0, \end{aligned}$$

where $(\tilde{c}_t, \tilde{n}_t)$ denote average per capita consumption and hours worked. The household takes these as given; they equal (c_t, n_t) in equilibrium. The household's

4.4 Application to DSGE Models

constraints are collectively

$$\begin{aligned}
 d_{t+1} &= (1 + r_t) d_t - \zeta_t + c_t + i_t + \frac{\phi}{2} (k_{t+1} - k_t)^2 \\
 \zeta_t &= A_t k_t^\alpha n_t^{1-\alpha} \\
 k_{t+1} &= \nu_t^{-1} i_t + (1 - \delta) k_t \\
 \ln A_{t+1} &= \rho_A \ln A_t + \varepsilon_{At+1} \\
 \ln r_{t+1} &= (1 - \rho_r) \ln \bar{r} + \rho_r \ln r_t + \varepsilon_{rt+1} \\
 \ln \nu_{t+1} &= \rho_\nu \ln \nu_t + \varepsilon_{\nu t+1} \\
 \ln \varphi_{t+1} &= \rho_\varphi \ln \varphi_t + \varepsilon_{\varphi t+1},
 \end{aligned}$$

where relative to the RBC model, the new variables are d_t , the stock of foreign debt, r_t , the exogenous interest rate at which domestic residents can borrow in international markets, ν_t , an investment-specific productivity shock, and the preference shock φ_t .

The state variables of the model are $(d_t, k_t, A_t, r_t, \nu_t, \varphi_t)$; the controls are (ζ_t, c_t, i_t, n_t) . In this application the non-linear policy functions

$$x_t = x(s_t), \quad x_t = (\zeta_t, c_t, i_t, n_t), \quad s_t = (d_t, k_t, A_t, r_t, \nu_t, \varphi_t)$$

were obtained using a second-order Taylor Series approximation of the system of expectational difference equations associated with the model, following Schmitt-Grohe and Uribe (2004). Given these policy functions, the state-transitions equations reduce to

$$d_{t+1} = (1 + r_t) d_t - \zeta(s_t) + c(s_t) + i(s_t) + \frac{\phi}{2} (k_{t+1} - k_t)^2 \quad (40)$$

$$k_{t+1} = \nu_t^{-1} i(s_t) + (1 - \delta) k_t \quad (41)$$

$$\ln A_{t+1} = \rho_A \ln A_t + \varepsilon_{At+1} \quad (42)$$

$$\ln r_{t+1} = (1 - \rho_r) \ln \bar{r} + \rho_r \ln r_t + \varepsilon_{rt+1} \quad (43)$$

$$\ln \nu_{t+1} = \rho_\nu \ln \nu_t + \varepsilon_{\nu t+1} \quad (44)$$

$$\ln \varphi_{t+1} = \rho_\varphi \ln \varphi_t + \varepsilon_{\varphi t+1}, \quad (45)$$

and the observation equations are

$$\ln(x_t/x(s_t)) = u_{x,t}, \quad x = \zeta, c, i, n, \quad (46)$$

$$u_{x,t} \sim N(0, \sigma_x^2). \quad (47)$$

4. NONLINEAR STATE-SPACE MODELS

As with the RBC model, hereafter state variables will be represented as logged deviations from steady state. In addition, given the form of (46), y_t is defined as $y_t = [\ln \zeta_t \quad \ln c_t \quad \ln i_t \quad \ln n_t]'$. All subsequent formulas should be read in accordance with these representations.

Notice that (40) and (41) characterize a bivariate degenerate transition of the form

$$q_t = \phi(p_{t-1}, q_{t-1}), \quad (48)$$

where following the notation of Section 4.3, $p_t = (A_t, r_t, v_t, \varphi)$, and $q_t = (d_t, k_t)$. Its inverse and corresponding linear approximation are denoted respectively as

$$q_{t-1} = \psi(q_t, p_{t-1}), \quad q_{t-1} = \tilde{\psi}(q_t, p_{t-1}).$$

The Jacobian associated with ψ is given by

$$J(q_t, p_{t-1}) = \left\| \frac{\partial}{\partial q_t'} \psi(q_t, p_{t-1}) \right\|. \quad (49)$$

The inversion of (40) and (41) is achieved as follows. It turns out that the transition equation for k_t is independent of d_t , thus it is possible to exploit the triangular structure of the system by first solving for k_{t-1} , and then using this result to find d_{t-1} . Defining $s_{t-1}^1 = [k_{t-1}, p_{t-1}]'$, the second-order approximation to the law of motion of k_t is given by

$$k_t = C_k + L_k s_{t-1}^1 + \frac{1}{2} s_{t-1}^{1'} Q_k s_{t-1}^1, \quad (50)$$

which is a quadratic equation in k_{t-1} with solutions

$$k_{t-1} = \frac{-b_k \pm \sqrt{b_k^2 - 4a_k c_k}}{2a_k}, \quad (51)$$

$$a_k = \frac{1}{2} Q_k^{11}, \quad (52)$$

$$b_k = L_k^1 + Q_k^{12} p_{t-1}, \quad (53)$$

$$c_k = C_k + L_k^2 p_{t-1} + \frac{1}{2} p_{t-1}' Q_k^{22} p_{t-1} - k_t, \quad (54)$$

$$Q_k = \begin{bmatrix} Q_k^{11} & Q_k^{12} \\ Q_k^{21} & Q_k^{22} \end{bmatrix}, \quad L_k = \begin{bmatrix} L_k^1 \\ L_k^2 \end{bmatrix}.$$

4.4 Application to DSGE Models

As the capital stock evolves slowly, the solution to (50) is chosen as

$$k_{t-1}^* = \arg \min [|(k_{t-1}^1 - k_t)|, |(k_{t-1}^2 - k_t)|], \quad (55)$$

where k_{t-1}^1 and k_{t-1}^2 are the roots (51).

Having obtained k_{t-1}^* , the solution of d_{t-1} proceeds as follows. Substituting (41) for k_{t+1} in (40), the second-order approximation to the law of motion of d_t is given by:

$$d_t = C_d + L_d s_{t-1} + \frac{1}{2} s_{t-1}' Q_d s_{t-1},$$

which is a quadratic equation in d_{t-1} with solutions

$$d_{t-1} = \frac{-b_d \pm \sqrt{b_d^2 - 4a_d c_d}}{2a_d},$$

$$a_d = \frac{1}{2} Q_d^{11}, \quad (56)$$

$$b_d = L_d^2 + Q_d^{12} s_{t-1}^1, \quad (57)$$

$$c_d = C_d + L_d^2 s_{t-1}^1 + \frac{1}{2} s_{t-1}^{1'} Q_k^{22} s_{t-1}^1 - d_t, \quad (58)$$

$$Q_d = \begin{bmatrix} Q_d^{11} & Q_d^{12} \\ Q_d^{21} & Q_d^{22} \end{bmatrix}, \quad L_d = \begin{bmatrix} L_d^1 \\ L_d^2 \end{bmatrix}.$$

Again the solution d_{t-1}^* is selected following (55).

The sequence of operations just described effectively transforms a triangular inverse transformation into a diagonal transformation. The corresponding Jacobian is given by

$$J(q_t, p_{t-1}) = (b_k^2 + 4a_k c_k)^{-\frac{1}{2}} \cdot (b_d^2 + 4a_d c_d)^{-\frac{1}{2}}.$$

Having achieved inversion, implementation of the EIS filter proceeds precisely as with the RBC model, with the following straightforward modifications:

- $J(k_t, z_{t-1})$ in (30) is replaced by $J(d_t, k_t, p_{t-1})$ in (49).
- The predictive density $f(s_t | Y_{t-1})$ in (34) becomes

$$f(s_t | Y_{t-1}) \simeq \int J(q_t, p_{t-1}) f(p_t | s_{t-1}) m(s_{t-1}; \widehat{a}_{t-1}) |_{q_{t-1}=\psi(q_t, p_{t-1})} dp_{t-1},$$

where

$$f(p_t | s_{t-1}) = N_4 \left(\begin{bmatrix} 0 & 0 & \rho_A & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho_r & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_v & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_\varphi \end{bmatrix} p_{t-1}, \begin{bmatrix} \sigma_A^2 & 0 & 0 & 0 \\ 0 & \sigma_r^2 & 0 & 0 \\ 0 & 0 & \sigma_v^2 & 0 \\ 0 & 0 & 0 & \sigma_\varphi^2 \end{bmatrix} \right).$$

4. NONLINEAR STATE-SPACE MODELS

- With y_t now defined as $y_t = [\ln \zeta_t \quad \ln c_t \quad \ln i_t \quad \ln n_t]'$, the measurement density becomes

$$f(y_t | s_t, Y_{t-1}) \sim N_3(\mu(s_t), V),$$

$$\mu(s_t) = \begin{bmatrix} \ln \zeta(s_t) \\ \ln c(s_t) \\ \ln i(s_t) \\ \ln n(s_t) \end{bmatrix}, \quad V = \begin{bmatrix} \sigma_y^2 & 0 & 0 & 0 \\ 0 & \sigma_c^2 & 0 & 0 \\ 0 & 0 & \sigma_i^2 & 0 \\ 0 & 0 & 0 & \sigma_n^2 \end{bmatrix}.$$

The performance of the EIS filter will be demonstrated with two Monte Carlo experiments patterned exactly after those used in working with the RBC model. Again working with two data sets: an artificial data set consisting of 100 realizations of $\{\zeta_t, c_t, i_t, n_t\}$ generated using the parameterization of the model given in Table 4.3; and a Canadian data set consisting of quarterly real per capita observations on $\{\zeta_t, c_t, i_t, n_t\}$, spanning 1976:I-2008:IV (132 observations), and detrended using the Hodrick-Prescott filter. The latter was obtained from Statistics Canada.

Aside from the parameters that characterize sources of stochastic uncertainty in the model, the artificial data were generated using the parameter values calibrated by Schmitt-Grohe and Uribe (2003) to match the summary statistics on Canadian data reported by Mendoza (1991): parameter values are listed in their Table 1 (and in Table 4.3 here), and the summary statistics in their Table 3. The parameters that characterize sources of stochastic uncertainty in the model were chosen as those that minimized the sum of squared differences between Mendoza's summary statistics (excluding the trade balance) and the statistics implied by the model; the statistics are standard deviations of $\{\zeta_t, c_t, i_t, n_t\}$, first-order serial correlations, and contemporaneous correlations with output. Finally, the standard deviations of all measurement errors were set at 0.5%. The same parameters used to generate the data were also used to evaluate the likelihood function in the MC experiment.

The parameters used to evaluate the likelihood function associated with the actual data are posterior modes estimated using the prior specification indicated in Table 4.3. The prior consists of independent normal distributions specified for each parameter. Aside from parameters that characterize stochastic uncertainty, prior means were set at the values specified by Schmitt-Grohe and Uribe, and prior standard deviations were set to reflect non-trivial uncertainty over these

4.4 Application to DSGE Models

Table 4.3: Parameter Values, SOE Model

	γ	ω	ψ	α	ϕ	\bar{r}	δ	ρ_A	σ_A	
Art. Data	2	1.455	0.11135	0.32	0.028	0.04	0.1	0.53	0.0089	
Prior Mean	2	1.455	0.11	0.32	0.028	0.007	0.025	0.8	0.005	
Prior Std. Dev.	1	0.2	0.001	0.05	0.01	0.025	0.025	0.2	0.005	
Post. Mode	2.49	1.33	0.11	0.23	0.039	0.02	0.02	0.82	0.0019	
	ρ_r	σ_r	ρ_v	σ_v	ρ_φ	σ_φ	σ_y	σ_c	σ_i	σ_n
Art. Data	0.37	0.001	0.89	0.001	0.3	0.0152	0.005	0.005	0.005	0.005
Prior Mean	0.8	0.0022	0.8	0.005	0.8	0.005	0.005	0.005	0.005	0.005
Prior Std. Dev.	0.2	0.0005	0.2	0.005	0.2	0.005	0.005	0.005	0.0005	0.005
Post. Mode	0.79	0.0022	0.87	0.001	0.86	0.0031	0.0038	0.0065	0.0046	0.0058

specifications. (Note that the specifications of δ and \bar{r} chosen by Schmitt-Grohe and Uribe are appropriate for annual data, and thus were translated into prior specifications appropriate for the quarterly observations employ here.) The priors over AR parameters were centered at 0.8 (s.d. 0.2); and with two exceptions along ill-behaved dimensions (σ_r and σ_i), the priors over σ 's were centered at 0.5% (s.d. 0.5%). The likelihood function implies strong negative correlation between σ_r and ρ_r , thus σ_r was set so that the posterior mode of ρ_r lied near its prior mean. Also, the posterior mode of σ_i was difficult to pin down, so its prior mean was centered at 0.5% like its counterparts, while its standard deviation was set to pin down the posterior mode at this value.

Results from the two MC experiments are presented in Table 4.4. Due to the increased dimensionality of the state space, N was set to 150,000 in working with the particle filter (requiring 128.59 and 169.66 seconds per function evaluation on a 963 GHz desktop computer using MATLAB for the artificial and real data sets respectively), and $N = R = 50$, $S = 30$ in working with the EIS filter (requiring 5.02 and 6.65 seconds per function evaluation).

To begin, it is interesting to stress that neither data set contains an outlier observation: in both data sets and across all variables, the largest deviation observed from sample means is 2.7 sample standard deviations. Despite this absence of outliers, there are significant differences between the likelihood values produced by the EIS and Kalman filters in both data sets. Relative to the RBC model, this reflects the added sources of non-linearity featured in the SOE model: e.g., the capital-adjustment cost term $\frac{\phi}{2} (k_{t+1} - k_t)^2$ in (40), and the endogenous discount factor θ_t featured in the household's objective function. Once again, when the

4. NONLINEAR STATE-SPACE MODELS

Table 4.4: Monte Carlo Means and Standard Deviations, SOE Model

	Particle Filter		EIS Filter		Kalman Filter
	Mean	Std. Dev.	Mean	Std. Dev.	
Artificial Data	1292.8274	2.0391	1283.6767	0.0721	1271.9214
Actual Data	1718.1382	0.4884	1713.6243	0.04771	1719.4209

EIS filter is implemented using linear model approximations, differences in log-likelihoods produced by the EIS and Kalman filters virtually disappear (becoming at most $6.3e-11$ across data sets and all time periods). Thus there is clearly a significant payoff in the implementation of a non-linear model representation in this application.

It is also interesting to note that in this application the artificial data set is the more challenging of the two. This is evident along two dimensions. First, MC standard deviations obtained using the artificial data set are relatively high for both filters. Second, the bias suffered by the particle filter is more substantial in the application involving the artificial data set. Specifically, the difference in mean log-likelihood approximations generated by the particle and EIS filters is more than 9 in working with the artificial data set, compared to less than 5 in working with the actual data set.

As opposed to the applications involving the RBC model, the explanation for these differences across data sets does not lie in the behavior of associated measurement errors: variances of measurement errors are closely comparable across data sets in this case. Instead, differences stem primarily from differences in the volatility and persistence of the model's structural shocks. In particular, with the model parameterization associated with the artificial data set calibrated to annual data, and the parameterization associated with the real data set estimated using quarterly observations, structural shocks are far less persistent, and generally more volatile, in the former case. The upshot is that in working with the actual data, the state variables are relatively easy to track, and in general the construction of likelihood approximations is less problematic.

Comparing the EIS and particle filters, as noted, the particle filter once again suffers non-trivial bias, on scales similar to those observed in working with the RBC model. Regarding MC standard errors, these differ by two orders of magnitude in the artificial data set, but by only one order of magnitude in the actual

data set. These results indicate that increases in the dimensionality of the state space do not necessarily amplify the numerical problems suffered by the particle filter: outliers and narrow measurement densities are far more important sources of difficulty.

To conclude the analysis of the SOE model, sampling errors associated with the log-likelihood estimates are reported in Table 4.4. Following the procedure described above in working with the RBC model, these errors are estimated to be 16.48 using the parameterization associated with the artificial data set, and 17.99 using the parameterization associated with the actual data set. Comparing these estimates with the MC standard errors reported in Table 4.4, it is possible to see that the particle filter serves as a better potential gauge of statistical uncertainty than was the case in the applications involving the RBC model. In particular, its MC standard errors are only 1/13th and 1/36th the size of their associated sampling errors in this case, while recall that in working with the RBC model, these ratios were roughly 2 and 1/92nd. The ratios associated with the EIS filter are 1/354th and 1/323rd in this case, compared with roughly 1/10,000 and 1/200 in working with the RBC model.

4.4.3 Repeated Samples

To this point the comparisons of the bias and numerical errors associated with the EIS and particle filters have been based on four data sets. This raises the question of whether the results reported are somehow sensitive to special features of these data sets. Thus a large set of additional experiments designed to address this issue were conducted.

Specifically, using each of the four models described above as data generating processes (the two parameterizations of the RBC and SOE models reported in Table 4.1 and 4.3), 100 artificial data sets of length T were generated (with T specified to match the corresponding data set associated with the parameterized model). For each realized data set, 100 sets of log-likelihood estimates using the EIS filter were obtained on a date-by-date basis using 100 sets of CRNs. For each data set, pseudo-true log-likelihood estimates using the EIS filter implemented by setting (N, R) as $(2000, 1000)$ were also obtained. Then the average (across CRNs) of the deviations from pseudo-true values of log-likelihood approximations

4. NONLINEAR STATE-SPACE MODELS

Table 4.5: EIS Filter Repeated Samples

DGP	Orig. DS	Bias		Orig. Smpl	Num. Acc.	
		Exp. Avg.	Exp. Stdv		Exp. Avg.	Exp. Stdv
RBC, Art.	0.1369	0.1364	0.0481	0.0015	0.0017	0.000589
RBC, Act.	0.8226	1.0619	0.0950	0.0044	0.0056	0.000383
SOE, Art.	0.5838	1.8279	1.3934	0.0053	0.0037	0.000786
SOE, Act.	0.8381	1.2202	0.1651	0.0026	0.0026	0.000173

Notes: Orig. DS denotes original data set; Exp. Avg. (Stdv) denotes average (standard deviation) of indicated statistic obtained across experimental data sets. RBC, Art. (Act.) denotes the parameterization of the RBC model associated with the artificial (actual) data set; likewise for the SOE model.

associated with the EIS filter for each data set and each time period were calculated. Standard deviation (over CRNs) of log-likelihood approximations were also computed. The former provides a measure of bias; the latter a measure of numerical accuracy.

Summing the bias measure over time periods, and averaging the numerical accuracy measure over time periods, a single measure of bias and numerical accuracy for each data set is obtained. Respectively, these measures are given by

$$\sum_{t=1}^T \frac{1}{M} \sum_{m=1}^M (l_{t,m} - l_t^*), \quad \frac{1}{T} \sum_{t=1}^T \sqrt{\frac{1}{M} \sum_{m=1}^M (l_{t,m} - \bar{l}_t^M)^2},$$

where $l_{t,m}$ denotes time- t log-likelihood calculated using the m^{th} set of CRNs, l_t^* denotes its pseudo-true value, \bar{l}_t^M denotes the sample average of $l_{t,m}$ obtained across CRNs, and M denotes the total number of CRNs obtained in the experiment. Assessing the mean and standard deviation of both measures calculated over artificial data sets, and comparing these with corresponding measures obtained using the four data sets analyzed above, we obtain context for interpreting the specific results reported above for the four original data sets. Results are reported in Table 4.5.

Note first that for the two cases in which the original data sets were generated artificially, bias and summary statistics lie within two standard deviations of the average values obtained in the repeated-sample experiments. This is as expected, since in these cases all 101 data sets (the original and those generated in the experiment) come from the same DGP.

4.4 Application to DSGE Models

Table 4.6: Repeated VAR Samples

DGP	Orig. Smpl	Bias		Orig. Smpl	Num. Acc.	
		Exp. Avg.	Exp. Stdv		Exp. Avg.	Exp. Stdv
RBC, Act.	0.8226	0.9105	0.0625	0.0044	0.0050	0.000350
SOE, Act.	0.8381	1.0341	0.1795	0.0026	0.0028	0.000158

Notes: Orig. Smpl denotes original sample; Exp. Avg. (Stdv) denotes average (standard deviation) of indicated statistic obtained across experimental data sets. RBC, Act. denotes a DGP based on an unrestricted vector autoregressions estimated using the actual data; likewise for the SOE model.

For the actual data associated with the SOE model, the measure of numerical accuracy exactly equals its corresponding experimental average, while the measure of bias is only 69% of its corresponding experimental average and lies 0.06 below the ± 2 standard deviation range around the average. For the actual data set associated with the RBC model, the measures of bias and numerical accuracy are only 77% and 79% of their corresponding experimental averages, and lie 0.05 and 0.0004 below their ± 2 standard deviation ranges. Given the two large outliers present in the original RBC data set, it is perhaps surprising that these measures are so close.

The differences noted for the actual data sets indicate that the theoretical models used as DGPs in these experiments do not produce data sets that closely mimic the actual series. To explore this possibility, the experiments was repeated using as alternative DGPs unrestricted vector autoregressions estimated using the original series. As reported in Table 4.6, in this case all summary statistics lied well within their associated ± 2 standard deviation ranges.

The bottom line that can be taken from these experiments is that the results detailed above for the four original data sets are largely representative of those one would expect to obtain in working with repeated samples from appropriate data generating processes. Moreover, the algorithm presented for implementing the EIS filter is remarkably reliable, having succeeded in quickly producing likelihood estimates for each of the many hundreds of data sets it confronted.

4.5 Conclusion

An efficient means of facilitating likelihood evaluation in applications involving non-linear and/or non-Gaussian state space representations was proposed: 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.

4.A Appendix 1: Nonlinear Approximate Solution of DSGE Models

Full information methods based on likelihood analysis have been used to estimate dynamic stochastic general equilibrium (DSGE) models since Sargent (1989). Since most DSGE models do not have analytical solution, and thus their likelihood function cannot be evaluated neither analytically nor numerically, most part of the literature have focused on (log-)linear approximations to the original models. This approach enables the use of Kalman filter to construct the likelihood function, and to perform estimation.

However, Fernandez-Villaverde et al. (2006) have proved that second-order approximation errors in the policy functions can have first order effects on the likelihood function. Moreover, they demonstrate that errors in the approximated likelihood accumulate as the sample size grows. As a result, likelihood approximation based on the linearized model can diverges from the exact one. Therefore, Fernandez-Villaverde and Rubio-Ramirez (2005, 2007), An and Schorfheide (2007), Amisano and Tristani (2007), and others have estimated nonlinear approximations to DSGE models in order to avoid first order errors in the approximated likelihood function, and report that the nonlinear approximations to these models deliver a better fit to the data. Since Kalman filter cannot be used to construct the likelihood function implied by nonlinear and/or nonnormal models, the estimation of such models are normally based on Sequential Monte Carlo algorithms.

4.A Appendix 1: Nonlinear Approximate Solution of DSGE Models

However, before one turns to the estimation of the model, it is necessary to decide upon which nonlinear solution method to be used. The properties of a given solution method are crucial to the estimation exercise, as the policy functions it provides are going to be used to compute the likelihood of the model. Furthermore, when estimating a DSGE model, it is interesting to have a solution method that is not only precise but also fast, as the likelihood function will need to be evaluated many times under different parameter values, implying that the model will have to be solved many times. The purpose of this appendix is to better understand the performance of different solution methods.

Aruoba et al. (2006) addressed this question in the context of an RBC model, while the focus of this appendix will be on the stochastic growth model exactly because the results for the RBC model are already known. Although the relative performance of different solution methods are model dependent, the fact that the results presented here are in line with those from Aruoba et al. (2006), signals that they might be robust across models.

The model is solved using second-order perturbation method from Schmitt-Grohe and Uribe (2004), which delivers local approximation to the policy functions, and using projection methods, that compute global approximate solutions. Two different projection methods are examined: the Chebyshev collocation method, and the finite elements method. For comparison purposes, a log-linear approximation is also computed and presented. Based on the policy function obtained, data from the different model approximations are simulated and their characteristics compared.

4.A-1 The Stochastic Growth Model

The model is of a representative household that seeks to maximize lifetime utility:

$$\max_{c_t} U = E_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{c_t}{1-\phi} \right)^{1-\phi},$$

where c_t is time- t consumption, β is a time-discount factor, and $1/\phi$ is the household's intertemporal elasticity of substitution.

4. NONLINEAR STATE-SPACE MODELS

The household's constraints are given by

$$y_t = z_t k_t^\alpha, \quad (4.A-1)$$

$$y_t = c_t + i_t, \quad (4.A-2)$$

$$k_{t+1} = i_t + (1 - \delta)k_t, \quad (4.A-3)$$

$$z_t = e^{\omega_t}, \quad \omega_t = \rho\omega_{t-1} + \varepsilon_t, \quad (4.A-4)$$

with (k_0, z_0) given. Here, y represents aggregate output, z total factor productivity (TFP), k the stock of physical capital, i investment, and ε an i.i.d. $\mathcal{N} \sim (0, \sigma^2)$ productivity innovation. Regarding parameters, α is capital's share of output, δ is the capital depreciation rate, and ρ determines the persistence of TFP innovations.

First-order conditions associated with the household's problem are given by

$$c_t^{-\phi} = \beta E_t \left\{ c_{t+1}^{-\phi} R_{t+1} \right\}, \quad (4.A-5)$$

where

$$R_{t+1} = \alpha z_{t+1} k_{t+1}^{\alpha-1} + 1 - \delta.$$

Coupled with the household's constraints, these conditions constitute the system of nonlinear stochastic difference equations that comprise the model

$$c_t^{-\phi} - \beta E_t \left\{ c_{t+1}^{-\phi} R_{t+1} \right\} = 0 \quad (4.A-6)$$

$$z_t k_t^\alpha - c_t - i_t = 0 \quad (4.A-7)$$

$$k_{t+1} - i_t - (1 - \delta)k_t = 0 \quad (4.A-8)$$

$$\omega_t - \rho\omega_{t-1} - \varepsilon_t = 0. \quad (4.A-9)$$

Steady state values \bar{x} for $x = \{y_t, c_t, i_t, k_t, z_t\}$ are derived by holding z_t to its steady state value \bar{z} , which is set to 1, and satisfy $c_t = \bar{c} \forall t$. From (4.A-5),

$$1 = \beta \left[\alpha \bar{k}^{\alpha-1} + 1 - \delta \right];$$

solving for \bar{k} :

$$(\beta\zeta)^{-1} = \left[\alpha \bar{k}^{\alpha-1} + 1 - \delta \right] \quad (4.A-10)$$

$$\frac{(\beta\zeta)^{-1} - 1 + \delta}{\alpha} = \bar{k}^{-(1-\alpha)} \quad (4.A-11)$$

$$\bar{k} = \left(\frac{\alpha}{(\beta\zeta)^{-1} - (1 - \delta)} \right)^{\frac{1}{1-\alpha}}. \quad (4.A-12)$$

4.A Appendix 1: Nonlinear Approximate Solution of DSGE Models

Then

$$\bar{y} = \bar{k}^\alpha, \quad (4.A-13)$$

$$\bar{i} = \left(\delta + \frac{g}{1-\alpha} \right) \bar{k} \quad (4.A-14)$$

$$\bar{c} = \bar{y} - \bar{i}. \quad (4.A-15)$$

4.A-2 Solution Methods

Solving for the equilibrium of the model presented above amounts to finding the policy functions that deliver the optimal choice of consumption as a function of contemporaneous technology shock and capital stock. As the system of equations (4.A-5)-(4.A-9) does not have a closed form solution, numerical methods have to be used to approximate the policy function characterizing its solution.

The most common method to solve dynamic equilibrium models is the perturbation method, which is based on Taylor series expansions of the agents' policy functions around the steady state of the economy and a perturbation parameter. When a first order Taylor series expansion is used, the model becomes linear in the state variables, allowing, for example, the estimation of the structural parameters using the Kalman filter. This approach was popularized under the name of (log-)linearization. Judd and Guu (1997) extended the method to compute higher-order terms of the expansion, and Schmitt-Grohe and Uribe (2004) and Kim et al. (2005) developed algorithms for the special case of second-order Taylor series expansions. Since perturbation approximations are based on Taylor series expansions, they are only valid locally, and provide poor approximations when the economy is far from the expansion point. This might be a problem if one is interested in analyzing economies subject to large shocks, as emerging markets, or extreme events like the Great Depression or the present financial crisis.

Global approximation methods are more general, and valid on the whole support of the approximated function. One class of such global methods are the projection methods (see DeJong and Dave, 2007; Judd, 1998; Mirana and Fackler, 2002). These methods are also known as weighted residual methods, as they take basis functions to build an approximated policy function that minimizes a residual function. Here two methods of this class are going to be used, one in which the basis functions are nonzero globally, the Chebyshev collocation method, and the finite elements method, whose basis functions are nonzero only locally.

4. NONLINEAR STATE-SPACE MODELS

Next the different solution methods are briefly presented, for a more thorough exposition of such methods see Heer and Maussner (2005), Judd (1998), or Mirana and Fackler (2002).

4.A-2.1 Perturbation

Perturbation methods take a problem that has a known solution for some parameter values and derive solutions for similar problems with nearby parameters. This approach relies on Taylor series expansions, on implicit function theorems, and on results from bifurcation theory. In applications involving DSGE models, approximations of the policy function are often constructed around the steady state and the perturbation parameter is usually chosen to be the standard deviation of the technology shock σ_ε .

Solving the stochastic growth model presented above reduces to solving

$$\begin{aligned} 0 &= E_t [f(c_{t+1}, c_t, s_{t+1}, s_t, \sigma_\varepsilon)] & (4.A-16) \\ 0 &= E_t \left[c_t^{-\phi} - \beta c_{t+1}^{-\phi} \left\{ \exp(\rho\omega_t + \varepsilon_{t+1})(z_t k_t^\alpha - c_t + (1-\delta)k_t)^{\alpha-1} + (1-\delta) \right\} \right], \end{aligned}$$

for a policy function

$$c_t = m(s_t, \sigma_\varepsilon), \quad (4.A-17)$$

where $s_t = [k_t, z_t]$ denotes the vector of state variables. The state transition equations can be written as

$$s_{t+1} = h(s_t, \sigma_\varepsilon) + \eta\sigma_\varepsilon\zeta, \quad (4.A-18)$$

where η is a vector with zero and ones defining the position of the state variable that is affected by the exogenous shock ζ , which in the stochastic growth model is normally distributed with mean zero and unit variance.

The perturbation method proceeds substituting the proposed solutions (4.A-17) and (4.A-18) into (4.A-16), yielding

$$\begin{aligned} F(s_t, \sigma_\varepsilon) &= E_t [f(g(h(s_t, \sigma_\varepsilon) + \eta\sigma_\varepsilon\zeta, \sigma_\varepsilon), m(s_t, \sigma_\varepsilon), h(s_t, \sigma_\varepsilon) + \eta\sigma_\varepsilon\zeta, s_t)] \\ &= 0. \end{aligned} \quad (4.A-19)$$

In general it is not possible to solve (4.A-16) for all values of the perturbation parameter σ_ε , but the steady state values in (4.A-12)-(4.A-15), per definition, solve the system for $\sigma_\varepsilon = 0$, thus serving as a benchmark from which solutions to

4.A Appendix 1: Nonlinear Approximate Solution of DSGE Models

nearby parameter values will be derived. If the interest is on first-order approximations, then

$$\begin{aligned} m(s_t, \sigma_\varepsilon) &= g(\bar{s}, 0) + g_s(\bar{s}, 0)(s_t - \bar{s}) + g_{\sigma_\varepsilon}(\bar{x}, 0)\sigma_\varepsilon \\ h(s_t, \sigma_\varepsilon) &= h(\bar{s}, 0) + h_s(\bar{s}, 0)(s_t - \bar{s}) + h_{\sigma_\varepsilon}(\bar{x}, 0)\sigma_\varepsilon, \end{aligned} \quad (4.A-20)$$

where $m_s(\bar{s}, 0)$ and $h_s(\bar{s}, 0)$ are vectors containing the first derivatives with respect to the states s_t , and are given by

$$g_s(\bar{s}, 0) = \begin{bmatrix} \frac{\partial g_s(\bar{s}, 0)}{\partial k_t} \\ \frac{\partial g_s(\bar{s}, 0)}{\partial z_t} \end{bmatrix}, \quad h_s(\bar{s}, 0) = \begin{bmatrix} \frac{\partial h_s(\bar{s}, 0)}{\partial k_t} \\ \frac{\partial h_s(\bar{s}, 0)}{\partial z_t} \end{bmatrix}.$$

By the definition of the non-stochastic steady state, $m(\bar{s}, 0) = \bar{c}$, and $h(\bar{s}, 0) = \bar{s}$. Furthermore, note that because $F(\bar{s}, 0)$ must be equal to zero, it must be the case that all its derivatives are also equal to zero. Thus, a system of quadratic equations on the unknown elements of g_x and h_x can be constructed by evaluating the derivatives of $F(s_t, \sigma_\varepsilon)$ at the steady state and $\sigma_\varepsilon = 0$. A number of authors have developed algorithms to find solutions to this problem that are associated with non-explosive paths for the states and controls (see Blanchard and Kahn, 1980; Klein, 2000; Sims, 2002). Similarly, using $F_{\sigma_\varepsilon}(\bar{s}, 0)$ it can be shown that g_{σ_ε} and h_{σ_ε} are equal to zero (see Schmitt-Grohe and Uribe, 2004).

A second-order approximation to (4.A-17) and (4.A-18) is obtained by adding second order terms to the expansion in (4.A-20):

$$\begin{aligned} m(s_t, \sigma_\varepsilon) &= g(\bar{s}, 0) + g_s(\bar{s}, 0)(s_t - \bar{s}) + \frac{1}{2}(s_t - \bar{s})' g_{ss}(\bar{s}, 0)(s_t - \bar{s}) + \frac{1}{2} g_{\sigma_\varepsilon \sigma_\varepsilon}(\bar{s}, 0) \sigma_\varepsilon^2 \\ h(s_t, \sigma_\varepsilon) &= h(\bar{s}, 0) + h_s(\bar{s}, 0)(s_t - \bar{s}) + \frac{1}{2}(s_t - \bar{s})' h_{ss}(\bar{s}, 0)(s_t - \bar{s}) + \frac{1}{2} h_{\sigma_\varepsilon \sigma_\varepsilon}(\bar{s}, 0) \sigma_\varepsilon^2, \end{aligned}$$

where

$$g_{ss}(\bar{s}, 0) = \begin{bmatrix} \frac{\partial^2 g_s(\bar{s}, 0)}{\partial k_t^2} & \frac{\partial^2 g_s(\bar{s}, 0)}{\partial z_t \partial k_t} \\ \frac{\partial^2 g_s(\bar{s}, 0)}{\partial k_t \partial z_t} & \frac{\partial^2 g_s(\bar{s}, 0)}{\partial z_t^2} \end{bmatrix}, \quad h_{ss}(\bar{s}, 0) = \begin{bmatrix} \frac{\partial^2 h_s(\bar{s}, 0)}{\partial k_t^2} & \frac{\partial^2 h_s(\bar{s}, 0)}{\partial z_t \partial k_t} \\ \frac{\partial^2 h_s(\bar{s}, 0)}{\partial k_t \partial z_t} & \frac{\partial^2 h_s(\bar{s}, 0)}{\partial z_t^2} \end{bmatrix}.$$

Plugging in the coefficients found in the first-order approximation it is possible to obtain a linear system in the second order terms of the policy function that is easily solvable. Taking higher order derivatives and plugging in the results found in the previous step, one is able to get higher order approximations.

4. NONLINEAR STATE-SPACE MODELS

4.A-2.2 Projection Methods

In Projection methods, the policy functions corresponding to the solution of the model are represented as a linear combination of known basis functions. The Weierstrass theorem presented in Section 2.1 suggests that polynomials might be a good choice for the elements of the basis for the function space of the approximation. The coefficients on each basis functions are the objects to be computed to obtain an approximate solution, and are found by setting a residual equation to zero on average.

For the stochastic growth model in (4.A-1)-(4.A-4), the problem is to find $c(s_t) : \mathbb{R}^2 \rightarrow \mathbb{R}$ that satisfies the functional equation (4.A-6). The approximation to the policy function will be of the form

$$\widehat{c}(s_t, \chi) = \sum_{i=1}^{n_k} \sum_{j=1}^{n_z} \chi_{i,j} P_{i,j}(s_t), \quad (4.A-21)$$

where $\chi_{i,j}$ are the unknown coefficients, n_k and n_z are the number of grid points in each state dimension, and $P_{i,j}(s)$ are the basis functions, which are often simple functions, like polynomials or piecewise linear functions.

For the stochastic growth, the residual equation is defined based on (4.A-16) as

$$Res(s_t; \chi) = F(\widehat{c}(s_t; \chi)), \quad (4.A-22)$$

where

$$F(\widehat{c}(s_t; \chi)) = E_t [\widehat{c}(s_t; \chi)^{-\phi} - \beta \widehat{c}(s_{t+1}(s_t, \varepsilon_{t+1}); \chi)^{-\phi} R(s_{t+1}(s_t, \varepsilon_{t+1}))] \quad (4.A-23)$$

and $s_{t+1}(s_t, \varepsilon_{t+1})$ represent the transition equation of the state variables (4.A-3)-(4.A-4), and the expectation is with respect to ε_{t+1} .

The idea of Projection methods is to chose χ_i so that the residual equation (4.A-22) can be close to zero in a weighted integral sense. That is, χ_i are chosen so that

$$\int w_i(s_t) Res(s_t; \chi) ds = 0, \quad i = 1, \dots, n, \quad (4.A-24)$$

where $n = n_k \cdot n_z$ are the total number of unknown coefficients χ , and $w_i(s_t), \forall i = 1, \dots, n$ are *weight functions* that can be different from $P_i(s_t)$. Therefore, instead

4.A Appendix 1: Nonlinear Approximate Solution of DSGE Models

of setting $Res(s_t; \chi)$ to zero for all s_t , the method sets a weighted integral of $Res(\cdot)$ to zero.

There are different ways to determine the coefficients χ_1, \dots, χ_n depending on the weight function used. The collocation method, which uses weights that are zero almost everywhere given by $w_i(s_t) = \delta(s_t - s_i)$ ¹, is going to be used together with spectral methods, where the basis functions $P_i(s_t)$ are nonzero almost everywhere. The Galerkin method uses $w_i(s_t) = P_i(s_t)$, forcing the residual to be orthogonal to each of the basis functions. This scheme is going to be used in the finite elements method, whose basis functions $P_i(s_t)$ are zero almost everywhere.

4.A-2.3 Finite Elements Method

The finite element method is an algorithm for solving functional equations widely used in engineering applications (see, e.g. Ern and Guermond, 2004). It subdivides the domain of the state space into nonintersecting subdomains (*elements*), and fits low-order polynomials on them instead of fitting high-order polynomials on the entire state space². This allows extra flexibility, as finer subdivision of the state space could be used in regions where nonlinearities occur, or higher order polynomials could be used where needed.

The objective is to find the vector of coefficients χ , which are used to represent the approximation to the consumption function $\widehat{c}(k_t, \theta_t, \chi) = \sum_{i=0}^n \chi_i P_i(s_t)$, and satisfy the residual function

$$Res(k_t, \theta_t; \chi) = \frac{\beta}{\sqrt{\pi}} \int \frac{\widehat{c}(k_{t+1}, \theta_{t+1}; \chi)^{-\phi}}{\widehat{c}(k_t, \theta_t; \chi)^{-\phi}} \left(\alpha k_{t+1}^{\alpha-1} \sqrt{\frac{1+\theta_{t+1}}{1-\theta_{t+1}}} + 1 - \delta \right) e^{-\nu^2} d\nu - 1, \quad (4.A-25)$$

where

$$\begin{aligned} k_{t+1}(s_t) &= k_t^\alpha \sqrt{\frac{1+\theta_t}{1-\theta_t}} + (1-\delta)k_t - \widehat{c}(s_t; \chi), \\ \theta_{t+1} &= \tanh \left(\rho \tanh^{-1}(\theta_t) + \sqrt{2}\sigma_\varepsilon \nu \right), \end{aligned} \quad (4.A-26)$$

and ν is normally distributed with mean 0 and variance 0.5. The upper bound to the capital stock is set to \bar{k} , thus the domain of the approximation $\widehat{c}(k_t, \theta_t, \chi)$

¹ $\delta(\cdot)$ denotes the Dirac delta function

²Spectral methods presented in the next subsection will adopt the opposite strategy of fitting high-order polynomials to the entire state space.

4. NONLINEAR STATE-SPACE MODELS

is defined as $\Omega = [0, \bar{k}] \times [-1, 1]$. The transformation in (4.A-26) allows the use of a variable that has compact support, θ , in place of a variable that does not, z . Furthermore, (4.A-26) uses the transformation introduced in equation 2.18 of Section 2.1.1.2 to compute expectations of Gaussian variables using the Gauss-Hermite formula. Since the integral in (4.A-25) is univariate and over a gaussian variable ν , Gauss-Hermite quadrature is a natural choice to be used to efficiently solve it. The application of the quadrature rule implies

$$R(k_t, \theta_t; \chi) \simeq \frac{\beta}{\sqrt{\pi}} \sum_{i=1}^n \frac{\widehat{c}(k_{t+1}, \theta_{t+1}^i; \chi)^{-\phi}}{\widehat{c}(k_t, \theta_t^i; \chi)^{-\phi}} \left(\alpha k_{t+1}^{\alpha-1} \sqrt{\frac{1 + \theta_{t+1}^i}{1 - \theta_{t+1}^i}} + 1 - \delta \right) w_i - 1,$$

where $\theta_{t+1}^i = \tanh(\rho \tanh^{-1}(\theta_t) + \sqrt{2}\sigma_\varepsilon \nu_i)$, and ν_i and w_i are the abscissas and weights for a N -point Gauss-Hermite quadrature rule.

Once the domain Ω of the approximation $\widehat{c}(k_t, \theta_t, \chi)$ is selected, one needs to divide it into non-overlapping elements $[k_i, k_{i+1}] \times [\theta_j, \theta_{j+1}]$ such that $\Omega = \bigcup_{i,j} [k_i, k_{i+1}] \times [\theta_j, \theta_{j+1}]$, where k_i is the i^{th} grid point for capital stock and θ_j the j^{th} grid point for the transformed TFP process³. In this application, smaller elements are placed where the policy function is steeper, and fewer larger elements where it is flatter.

The next step is to decide upon the interpolation functions that are going to be used to approximate the policy function in each element. Although quadratic, cubic and other polynomials functions can be used, linear basis provide a good approximation for this application. Set $P_{i,j}(s) = P_i(k)P_j(\theta)$, where

$$P_i(k) = \begin{cases} \frac{k-k_{i-1}}{k_i-k_{i-1}} & \text{if } k \in [k_{i-1}, k_i], \\ \frac{k_{i+1}-k}{k_{i+1}-k_i} & \text{if } k \in [k_i, k_{i+1}], \\ 0 & \text{elsewhere,} \end{cases} \quad P_j(\theta) = \begin{cases} \frac{\theta-\theta_{j-1}}{\theta_j-\theta_{j-1}} & \text{if } \theta \in [\theta_{j-1}, \theta_j], \\ \frac{\theta_{j+1}-\theta}{\theta_{j+1}-\theta_j} & \text{if } \theta \in [\theta_j, \theta_{j+1}], \\ 0 & \text{elsewhere,} \end{cases}$$

are the interpolation functions. The function $P_{i,j}(s)$ is 0 everywhere, except inside four elements.

The unknown coefficients $\chi_{i,j}$ are found using the Galerkin method, which implies that the weighting functions $w_{i,j}$ are chosen to be the basis functions $P_{i,j}(s)$. Using condition (4.A-24), $\chi_{i,j}$ is the solution to the following system of

³Rectangular elements are used here, but other polygons or polyhedrons can be utilized (see Ern and Guermond, 2004).

4.A Appendix 1: Nonlinear Approximate Solution of DSGE Models

equation

$$\int_{[k_{i-1}, k_i] \times [\theta_{j-1}, \theta_j] \cup [k_i, k_{i+1}] \times [\theta_j, \theta_{j+1}]} P_{i,j}(k, \theta) Res(k, \theta; \chi) dk d\theta = 0 \quad \forall i, j. \quad (4.A-27)$$

The integral in (4.A-27) can be evaluated using Gauss-Legendre quadrature (see Section 2.1.1.2, or Press et al. (2007)). Using 93 unequal elements in the capital stock dimension and 51 in the TFP dimension, 3 nodes for the Gauss-Legendre integration, and 7 nodes for the Gauss-Hermite integration implies that (4.A-27) is a system of 4743 nonlinear equations, which is solved efficiently using a Newton algorithm and exploiting the sparsity of the system⁴.

4.A-2.4 Spectral Method

Spectral methods (see Judd, 1998) uses basis functions that are nonzero almost everywhere. One example of such bases are orthogonal polynomials (see Theorem 2.1.3) such as Chebyshev polynomials. Chebyshev polynomial interpolants are nearly optimal polynomial approximants, and the error associated with a n -degree Chebyshev polynomial interpolant cannot be larger than $2\pi \log(n) + 2$ times the lowest error attainable with any other polynomial approximant of the same order (see Mirana and Fackler, 2002).

A n^{th} order Chebyshev polynomial can be defined recursively as

$$T_0(s) = 1, T_1(s) = \tilde{s}, T_2(s) = 2\tilde{s}T_1(s) - T_0(s), \dots, T_n(s) = 2\tilde{s}T_{n-1}(s) - T_{n-2}(s),$$

where $\tilde{s} \in [-1, 1]$, thus the following transformation is used to map a state variable defined over a range $[\underline{s}, \bar{s}]$ into $[-1, 1]$:

$$\tilde{s} = \frac{s - s^*}{\vartheta}, \quad (4.A-28)$$

where s^* are the steady state values of s , and $\vartheta = \bar{s} - s^* = s^* - \underline{s}$. The approximation to the policy function is then given by

$$\tilde{c}(s, \chi) = \sum_{i=1}^{n_k} \sum_{j=1}^{n_z} \chi_{i,j} T_{i,j}(s), \quad (4.A-29)$$

⁴Analytical derivatives of the residual function with respect to $\chi_{i,j}$ are calculated to speed up computations as in McGrattan (2006).

4. NONLINEAR STATE-SPACE MODELS

Rewriting the residual equation (4.A-22) using the Chebyshev approximation to the policy function yields

$$Res(s_t; \chi) = \frac{\beta}{\sqrt{2\pi\sigma_\varepsilon}} \int \left[\frac{\widehat{c}(s_t; \chi)^{-\tau}}{\widehat{c}(s_{t+1}; \chi)^{-\tau}} R(s_{t+1}) \right] \exp\left(\frac{\varepsilon_{t+1}^2}{2\sigma_\varepsilon^2}\right) d\varepsilon, \quad (4.A-30)$$

where s_{t+1} are given by the law of motion of capital (4.A-3), and of TFP (4.A-4). The approximation of the univariate integral in (4.A-30) is achieved via a Gauss-Hermite quadrature approximation.

Instead of using the Galerkin method, where the weight functions are given by the basis functions, a collocation weighting scheme is used. The *collocation nodes* are the weighting function of the collocation method, and for the case of Chebyshev polynomials basis functions the collocation nodes are often set to be the zeros of the Chebyshev polynomial selected. This choice ensures that the approximation error converges to zero as the degree of the polynomial (and thus the number of nodes) increases (see Judd, 1998). Because the weighting function is zero almost everywhere, the integral in (4.A-24) is avoided and χ is found by solving the residual function (4.A-30) at the collocation nodes.

Using 18th order polynomial for both state variables, the problem becomes one of solving the following system of 324 nonlinear equations

$$Res(k_i, z_j; \chi) = 0 \quad \forall i, j, \quad (4.A-31)$$

which is solved using a quasi-Newton method. Starting values used to initiate the algorithm are constructed using a linear approximation. Specifically, the log-linearized consumption policy function has the following form

$$\check{c}_t = F_{c,k} \check{k}_t + F_{c,z} \check{z}_t,$$

where $F_{c,s}$ is the elasticity of c_t with respect to the states, and \check{x} denotes log deviations from the steady-state of x . In terms of levels, an expanded Taylor Series approximation is given by

$$\begin{aligned} \widehat{c} &\approx \bar{c} + \frac{\bar{c}}{\bar{k}} F_{c,k} (k - \bar{k}) + \frac{\bar{c}}{\bar{z}} F_{c,z} (z - \bar{z}) \\ &\quad + \frac{1}{2} \left(\frac{\bar{c}}{\bar{k}} F_{c,k} \right) \left(\frac{\bar{c}}{\bar{z}} F_{c,z} \right) (k - \bar{k})(z - \bar{z}). \end{aligned}$$

4.A Appendix 1: Nonlinear Approximate Solution of DSGE Models

In turn, the approximation based on Chebyshev polynomials is of the form

$$\begin{aligned}\widehat{c} \approx & \chi_{11} + \chi_{12} \left(\frac{k_t - \bar{k}}{\vartheta_k} \right) + \chi_{21} \left(\frac{z_t - \bar{z}}{\vartheta_z} \right) \\ & + \chi_{22} \left(\frac{k_t - \bar{k}}{\vartheta_k} \right) \left(\frac{z_t - \bar{z}}{\vartheta_z} \right) + \dots\end{aligned}$$

Matching terms yields starting values

$$\begin{aligned}\chi_{11} &= \bar{c}, \quad \chi_{12} = F_{c,k} \vartheta_k \frac{\bar{c}}{\bar{k}}, \quad \chi_{21} = F_{c,z} \vartheta_z \frac{\bar{c}}{\bar{z}}, \\ \chi_{22} &= \frac{1}{2} \left(F_{c,k} \vartheta_k \frac{\bar{c}}{\bar{k}} \right) \left(\sigma \vartheta_z \frac{\bar{c}}{\bar{z}} \right).\end{aligned}$$

4.A-3 Numerical Comparisons

Two sets of parameter values are used to compare the different solution methods: a benchmark calibration, and an extreme calibration where shocks to the TFP process are more than 10 time more volatile. The objective is to see how the different solution methods behave as the shocks to the model economy become larger, pushing it away from the steady-state. More specifically, it is interesting to know how perturbation methods, which are based on local Taylor-series approximation, behave as the model drifts away from the expansion point. The two parameterizations used are given in Table 4.A-1.

As economists often use statistics from time series generated by the model economy to draw their conclusions, simulations of 600 observations each are computed for all solution methods using the same set of CRN to make them comparable. Artificial time series for consumption and capital stock are generated and the comparisons focus on their empirical moments, the empirical distribution, and on the policy function generated by each method.

Table 4.A-2 present the empirical mean, standard deviation, and first-order autocorrelation of the time series generated by the competing methods. All four

Table 4.A-1: Parameter Values, Stochastic Growth Model

	α	β	ϕ	δ	ρ	σ_ε
<i>Benchmark</i>	0.33	0.9896	2	0.0196	0.95	0.007
<i>Extreme</i>	0.33	0.9896	2	0.0196	0.95	0.1

4. NONLINEAR STATE-SPACE MODELS

Table 4.A-2: Empirical Moments, Benchmark Calibration

	Consumption			Capital Stock		
	μ	σ	ρ	μ	σ	ρ
Chebyshev	2.5546	0.0547	0.9931	35.6925	1.3052	0.9986
Finite Elements	2.5546	0.0547	0.9930	35.6922	1.3048	0.9986
2 nd Order	2.5546	0.0547	0.9931	35.6926	1.3067	0.9986
Log-Linear	2.5539	0.0547	0.9931	35.6637	1.3056	0.9986

μ denotes the empirical mean, σ the empirical standard error, and ρ the first-order autocorrelation.

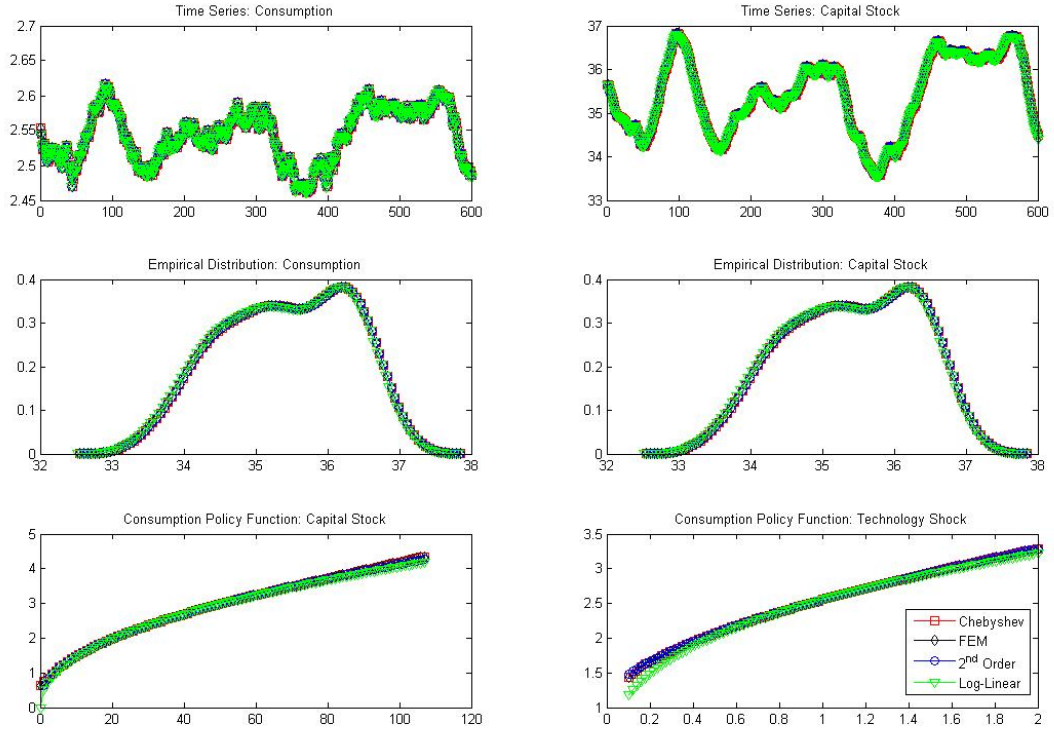
solution methods deliver very similar results, as the model calibrated using the benchmark parameters is almost linear. The two graphs on top of Figure 4.A-1 present artificial time series for consumption and capital stock generated by the different approximations, the pair of graphs in the middle are the empirical distribution of these two variables, and the lower graphs display the different approximations to the policy function. The different answers are almost indistinguishable, except for the log-linear policy rules for values of the state variables away from their steady state. In the lower right panel of Figure 4.A-1 it is clear how the log-linear approximation is pushed away from the others for low and high values for the TFP and for the capital stock. Although the second order perturbation method is only a local approximation, its policy functions are similar to the global ones for a very large range of the state space.

Once the standard deviation of the shock is increased, larger differences appear and the local methods deviate from the global ones. Table 4.A-3 contain the empirical moments of the artificial time series generated by approximations based on the extreme calibration. The global methods provide very similar results, which are quite different from the ones given by the log-linear approximation. Although the first-order autocorrelation for consumption and capital stock are very similar across methods, the log-linear empirical means and standard deviations are much lower. The second order perturbation method provide results that are closer to the ones of the global methods, but smaller differences in the empirical mean, and larger differences in the standard deviation of capital stock can be observed.

The time series plots in Figure 4.A-2 clearly show how the series based on

4.A Appendix 1: Nonlinear Approximate Solution of DSGE Models

Figure 4.A-1: Benchmark Calibration

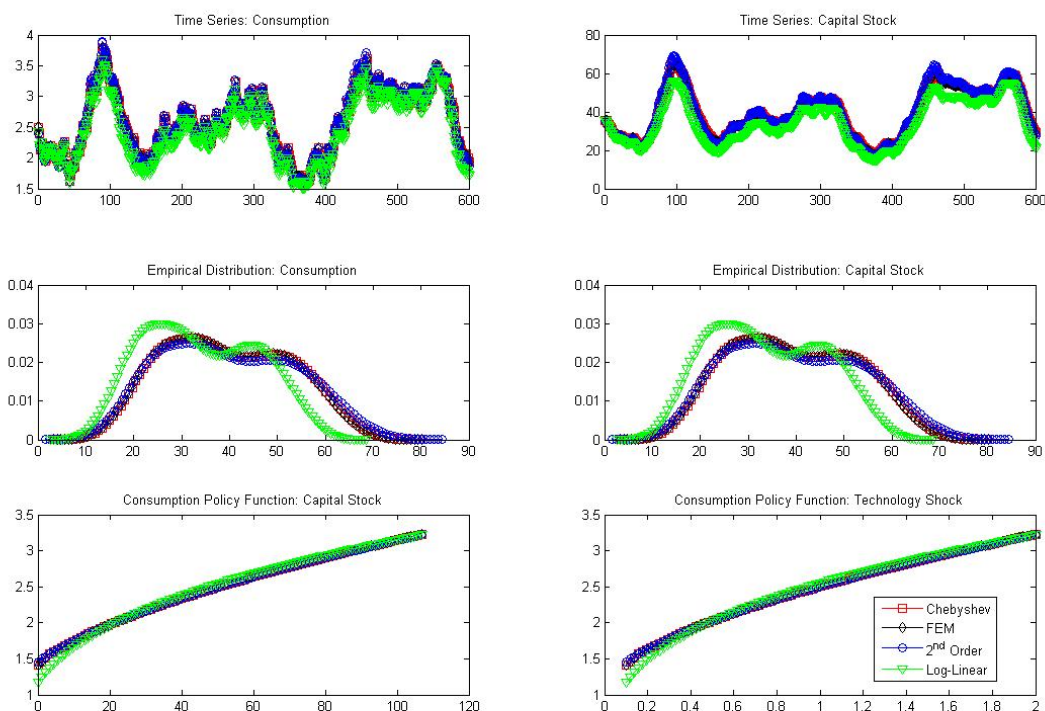


the log-linear approximation are always below the other series, evidencing the first order errors implied by the log-linear solution. Furthermore, as a result of this downward bias, the empirical distribution of consumption and capital for the extreme calibration is shifted to the left. The empirical distribution based on the second order approximation is much closer to the ones of the global methods, but it is possible to see how the right tails decay much slower than the ones of the distributions implied by the Chebyshev and finite elements methods, which is responsible for the larger standard deviations presented in Table 4.A-3. The policy functions of the three nonlinear approximations are again almost indistinguishable, but now the one based on the log-linear solution clearly deviate from the others in almost all the support of the state space, and this deviation becomes larger for low values of capital stock and TFP.

Table 4.A-4 present computing time for each of the methods analyzed. At least

4. NONLINEAR STATE-SPACE MODELS

Figure 4.A-2: Extreme Calibration



for the two-states stochastic growth model used here, Chebyshev polynomials and the collocation method seem to deliver a good balance between computing time and accuracy, although this method is known to suffer badly from the *curse of dimensionality*. The finite elements method approximations seem to be as good as the ones based on Chebyshev polynomials, but its computing time is much higher. The computing time of the finite elements could be improved via the development of more efficient mesh, but this would imply a much higher implementation time. The 2nd order perturbation method is the most straightforward nonlinear method to implement, and is also the fastest, but its approximation error grows with the nonlinearity of the model.

The numerical results for both calibrations of the stochastic growth model show that log-linearization is a good approximation method only when the model is relatively close to the steady-state. Second order approximations on the other

4.A Appendix 1: Nonlinear Approximate Solution of DSGE Models

Table 4.A-3: Empirical Moments, Extreme Calibration

	Consumption			Capital Stock		
	μ	σ	ρ	μ	σ	ρ
Chebyshev	2.8048	0.8499	0.9929	46.5106	23.2938	0.9983
Finite Elements	2.8091	0.8499	0.9929	46.5149	23.2938	0.9983
2 nd Order	2.8220	0.8675	0.9930	48.2093	26.7497	0.9984
Log-Linear	2.6744	0.8366	0.9927	40.8231	22.9544	0.9985

μ denotes the empirical mean, σ the empirical standard error, and ρ the first-order autocorrelation.

Table 4.A-4: Computational Time

	Benchmark Calibration	Extreme Calibration
<i>Chebyshev</i>	0.4907	0.9585
<i>FiniteElements</i>	10.5921	10.7950
<i>2ndOrder</i>	0.3521	0.3521
<i>Log – Linear</i>	0.0081	0.0081

Time in seconds computed using MATLAB version 7.8 on a 1.7 GHz desktop with 3.3GB of RAM memory.

4. NONLINEAR STATE-SPACE MODELS

hand, are able to approximate the solution to the stochastic growth model fairly well even when the model is subject to larger shocks that drag the economy away from its steady state. Although, this approach generates data that are more volatile than the one generated by the global methods analyzed in the extreme calibration case. Global methods appear to be the best approximation method, but their implementation is much more complicated than that of perturbation methods, and they are normally slower than the other competing methods. Perturbation methods seem to deliver a good trade-off between programming and computational time, and approximation error.

Chapter 5

Conclusions

This thesis has focused on econometric applications requiring multivariate numerical integration. Models that attempt to capture real-world complexities are typically nonlinear and display many unobservable factors. These characteristics imply that the likelihood function of these models contain high-dimensional integrals that cannot be solved analytically, and have to be approximated numerically.

Importance Sampling is an important Monte Carlo simulation method often used to solve high-dimensional integrals, but given the difficulty in finding mechanical multivariate importance samplers, this method have not experienced the same success of other simulation approaches like MCMC methods (van Dijk, 1999). Here the Efficient Importance Sampling method developed by Richard and Zhang (2007) was used to overcome the problem of finding multivariate importance samplers with remarkable success. It was shown how importance sampling can be used to efficiently solve high dimensional integration problems in econometric models using panel data, and time series.

In chapter 3, EIS was used to estimate different non-linear panel data specifications in order to investigate the causes and dynamics of current account reversals in low- and middle-income countries. In particular, four sources of serial persistence were analyzed: (i) a country-specific random effect reflecting time-invariant differences in institutional, political or economic factors; (ii) serially correlated transitory error component capturing persistent country-specific shocks; (iii) dynamic common time-specific factor, designed to account for potential spill-over effects and global shocks to all countries; and (iv) a state dependence

5. CONCLUSIONS

component to control for the effect of previous events of current account reversal and to capture slow adjustments in international trade flows.

The likelihood evaluation of panel models with country-specific random heterogeneity require univariate integrals that were efficiently solved using Gauss-Hermite quadrature. For likelihood-based estimation of the panel model with country-specific random heterogeneity and serially correlated error components, the EIS was used to perform high-dimensional integration. The application of EIS allows for numerically very accurate and reliable ML estimation of this model. In particular, it improves significantly the numerical efficiency of GHK, which is the most frequently used MC procedure to estimate non-linear panel models with serially correlated errors.

The empirical results indicate that the static pooled probit model is strongly dominated by the alternative models with serial dependence. However, state-dependence and transitory country-specific errors are essentially observationally equivalent. Both sources of serial dependence are jointly significant only with the inclusion of random time-specific effects into the model with state-dependence, even though the time-specific effect is small with limited effect on the overall fit of the model. Also, conclusive evidence for the existence of random country-specific effects was not found.

Overall, the results relative to the determinants of current account reversals are in line with the those in the empirical literature on current account crises and confirm the empirical relevance of theoretical solvency and sustainability considerations with respect to a country's trade balance. More specifically, countries with high current account imbalances, low foreign reserves, a small fraction of concessional debt, and unfavorable terms of trades are more likely to experience a current account reversal. These results are fairly robust against the dynamic specification of the model.

Chapter 4 focus on likelihood evaluation of nonlinear DSGE models. Recently, a growing literature has estimated nonlinear DSGE model using the particle filter (see Amisano and Tristani, 2007; An and Schorfheide, 2007; Fernandez-Villaverde and Rubio-Ramirez, 2005, 2007), but little attention has been given to the extra uncertainty introduced via the sequential Monte Carlo filter. The particle filter is conceptually simple and easy to program, but it may suffer from severe sample

impoverishment and likelihood discontinuities given its discrete and fixed-support approximations to unknown densities in the filtering process.

An efficient means of facilitating likelihood evaluation in applications involving non-linear and/or non-Gaussian state space representations was presented: the EIS filter. The filter is adapted using the EIS 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 was shown to be substantial.

Two nonlinear DSGE models were used to illustrate the problems of likelihood evaluation using the particle filter, to explain the implementation of the EIS filter, and to show how it is able to overcome the problems. The first one is a standard two-state real business cycle (RBC) model estimated using the particle filter by Fernandez-Villaverde and Rubio-Ramirez (2005); the second is a small-open-economy (SOE) model patterned after those considered, e.g., by Mendoza (1991) and Schmitt-Grohe and Uribe (2003), but extended to include six state variables.

Both data sets used in the likelihood evaluation of the RBC model were used by Fernandez-Villaverde and Rubio-Ramirez (2005), and each of them poses a distinct challenge to efficient filtering. In the artificial data set, the standard deviations of the measurement errors are small relative to shocks to the unobserved states, which led to problems associated with sample impoverishment. In the real data set, the investment series contains two outliers, which induce non-trivial bias in likelihood estimates associated with the particle filter. Both of these challenges were overcome via implementation of the EIS filter.

The six-state variables small-open-economy model showed that the EIS filter extraordinary success in resolving the problems of the particle filter is not restricted to small state space systems. In this application neither data set contains an outlier observation nor standard deviations of measurement errors are small. Differences stem primarily from differences in the volatility and persistence of the model's structural shocks. In particular, when structural shocks are more persistent, and less volatile, the state variables are relatively easy to track, and in general the construction of likelihood approximations is less problematic. The particle filter once again suffers non-trivial bias on scales similar to those

5. CONCLUSIONS

observed in working with the RBC model even in a less problematic setup. The MC standard errors of the particle filter and EIS filter differ by two orders of magnitude in the artificial data set, and by one order of magnitude in the actual data set.

The Appendix to Chapter 4 presented a numerical comparison from different methods used in the literature to solve nonlinear DSGE models. Numerical results were based on two different calibrations of a stochastic growth model: a benchmark one; and an extreme calibration where the standard deviation of the shock to the total factor productivity is 10 times larger.

For both calibrations the log-linearization is a good approximation method only when the model is relatively close to the steady-state. Second order approximations on the other hand, are able to approximate the solution to the stochastic growth model fairly well even when the model is subject to larger shocks that drag the economy away from its steady state. Although, this solution approach generates data that are more volatile than the one generated by the global methods in the case of the extreme calibration. Global methods appear to be the best approximation method, but their implementation is much more complicated than that of perturbation methods, and they are normally slower than the competing methods. Therefore, perturbation methods seem to deliver a good trade-off between programming and computational time, and approximation error.

References

- Acton, F. S. (1997). *Numerical Methods that Work*. Cambridge University Press. 27
- Albert, J. H. and Chib, S. (1993). Bayes inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts. *Journal of Business & Economic Statistics*, 11(1):1–15. 16
- Amisano, G. and Tristani, O. (2007). Euro area inflation persistence in an estimated nonlinear dsge model. CEPR Discussion Papers 6373, C.E.P.R. Discussion Papers. 22, 148, 166
- An, S. and Schorfheide, F. (2007). Bayesian analysis of dsge models. *Econometric Reviews*, 26:113–172. 22, 148, 166
- Aruoba, S. B., Fernandez-Villaverde, J., and Rubio-Ramirez, J. (2006). Comparing solution methods for dynamic equilibrium economies. *Journal of Economic Dynamics and Control*, 30(12):2477–2508. 14, 23, 149
- Axtell, R. L. (2000). Why agents? on the varied motivations for agent computing in the social sciences. In *Working Paper 17, Center on Social and Economic Dynamics, Brookings Institution*, volume 17. 14
- Baldwin, R. and Krugman, P. (1989). Persistent trade effects of large exchange rate shocks. *The Quarterly Journal of Economics*, 104(4):635–54. 68
- Baltagi, B. H. (2005). *Econometric Analysis of Panel Data*. John Wiley & Sons. 69
- Bellman, R. (1961). *Adaptive Control Processes: A Guided Tour*. Princeton University Press. 26

REFERENCES

- Billio, M. and Monfort, A. (1998). Switching state-space models likelihood function, filtering and smoothing. *Journal of Statistical Planning and Inference*, 68(1):65–103. 15, 16
- Blanchard, O. J. and Kahn, C. M. (1980). The solution of linear difference models under rational expectations. *Econometrica*, 48(5):1305–11. 13, 153
- Bungartz, H.-J. and Dirnstorfer, S. (2003). Multivariate quadrature on adaptive sparse grids. *Computing*, 71. 34
- Butler, J. S. and Moffitt, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica*, 50(3):761–64. 20, 67
- Carpenter, J., Clifford, P., and Fearnhead, P. (1999). An improved particle filter for non-linear problems. *IEE Proceedings-Radar, Sonar and Navigation*, 146(1):2–7. 110
- Chabert, J.-L. (1999). *A History of Algorithms: From the Pebble to the Microchip*. Springer-Verlag. 29
- Chib, S. (2001). Markov chain monte carlo methods: computation and inference. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 5 of *Handbook of Econometrics*, chapter 57, pages 3569–3649. Elsevier. 17
- Corsetti, G., Pesenti, P., Roubini, N., and Tille, C. (1999). Competitive devaluations: A welfare-based approach. NBER Working Papers 6889, National Bureau of Economic Research, Inc. 64, 68, 69
- Creel, M. (2008). Estimation of dynamic latent variable models using simulated nonparametric moments. UFAE and IAE Working Papers 725.08, Unitat de Fonaments de l'Anàlisi Econòmica (UAB) and Institut d'Anàlisi Econòmica (CSIC). 15
- Davis, P. J. (1963). *Interpolation and Approximation*. Waltham Press. 26, 28
- Davis, P. J. and Rabinowitz, P. (1984). *Methods of Numerical Integration*. New York Academic Press. 19, 26, 31

REFERENCES

- DeGroot, M. H. (2004). *Optimal Statistical Decisions (Wiley Classics Library)*. Wiley-Interscience. 75, 117
- DeJong, D. N. and Dave, C. (2007). *Structural Macroeconometrics*. Princeton University Press. 119, 151
- DeJong, D. N., Hariharan, D., Liesenfeld, R., and Richard, J.-F. (2008). Efficient filtering in state-space representations. 113
- DeJong, D. N., Ingram, B. F., and Whiteman, C. H. (2000). A bayesian approach to dynamic macroeconomics. *Journal of Econometrics*, 98(2):203–223. 107
- Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag. 39
- Dixit, A. (1992). Investment and hysteresis. *Journal of Economic Perspectives*, 6(1):107–32. 68
- Dornbusch, R., Park, Y. C., and Claessens, S. (2000). Contagion: Understanding how it spreads. *World Bank Research Observer*, 15(2):177–97. 62, 69
- Doucet, A., de Freitas, N., and Gordon, N., editors (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag. 22, 105, 106
- Doucet, A. and Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. In Crisan, D. and Rozovsky, B., editors, *Handbook of Nonlinear Filtering*. Oxford University Press. 44, 45
- Duffie, D. and Singleton, K. J. (1993). Simulated moments estimation of markov models of asset prices. *Econometrica*, 61(4):929–52. 16
- Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press. 105
- Eckhardt, R. (1987). Stan ulam, john von neumann, and the monte carlo method. *Los Alamos Science*, pages 131–143. 35
- Edwards, S. (2004a). Financial openness, sudden stops, and current-account reversals. *American Economic Review*, 94(2):59–64. 60, 67, 80

REFERENCES

- Edwards, S. (2004b). Thirty years of current account imbalances, current account reversals and sudden stops. Technical report. 60, 67, 80
- Edwards, S. and Rigobon, R. (2002). Currency crises and contagion: an introduction. *Journal of Development Economics*, 69(2):307–313. 64, 68
- Ern, A. and Guermond, J.-L. (2004). *Theory and Practice of Finite Elements*. Graduate Texts in Mathematics. Springer, New York, Berlin, Heidelberg. 155, 156
- Falcetti, E. and Tudela, M. (2006). Modelling currency crises in emerging markets: A dynamic probit model with unobserved heterogeneity and autocorrelated errors. *Oxford Bulletin of Economics and Statistics*, 68(4):445–471. 21, 59, 61
- Fernandez-Villaverde, J. and Rubio-Ramirez, J. (2005). Estimating dynamic equilibrium economies: Linear versus nonlinear likelihood. *Journal of Applied Econometrics*, 20(7):891–910. 15, 22, 23, 105, 107, 118, 124, 148, 166, 167
- Fernandez-Villaverde, J. and Rubio-Ramirez, J. (2007). Estimating macroeconomic model: A likelihood approach. *Mimeo, University of Pennsylvania, ...* 22, 106, 107, 118, 148, 166
- Fernandez-Villaverde, J., Rubio-Ramirez, J. F., and Santos, M. S. (2006). Convergence properties of the likelihood of computed dynamic models. *Econometrica*, 74(1):93–119. 148
- Fisher, S. (1998). The imf and the asian crisis. Forum funds lecture at ucla on march 20 1998, International Monetary Fund. 21, 69
- Fishman, G. S. (1996). *Monte Carlo: Concepts, Algorithms, and Applications*. Springer-Verlag. 20, 26, 39
- Geweke, J. (1989a). Bayesian inference in econometric models using monte carlo integration. *Econometrica*, 57(6):1317–1339. 16, 39, 41, 44, 114, 115
- Geweke, J. (1989b). Efficient simulation from the multivariate normal distribution subject to linear inequality constraints and the evolution of constraint probabilities. Discussion paper, Duke University. 46

REFERENCES

- Geweke, J. (1996). Monte carlo simulation and numerical integration. In Amman, H. M., Kendrick, D. A., and Rust, J., editors, *Handbook of Computational Economics, vol. I*. Elsevier. 34, 39
- Geweke, J. (1999). Using simulation methods for bayesian econometric models: inference, development, and communication. *Econometric Reviews*, 18(1):1–73. 18
- Geweke, J. and Keane, M. (2001). Computationally intensive methods for integration in econometrics. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 5 of *Handbook of Econometrics*, chapter 56, pages 3463–3568. Elsevier. 21, 59
- Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2):107–113. 105, 109, 110
- Gourieroux, C. and Monfort, A. (2002). *Simulation-Based Econometric Methods*. Oxford University Press. 14, 45
- Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8(S):S85–118. 16
- Greene, W. (2004). Convenient estimators for the panel probit model: Further results. *Empirical Economics*, 29(1):21–47. 59
- Greene, W. H. (2007). *Econometric Analysis*. Prentice Hall, 6th edition. 68
- Haber, S. (1970). Numerical evaluation of multiple integrals. *SIAM Reviews*, 12(4):481–526. 31, 35
- Hajivassiliou, V., McFadden, D., and Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of Econometrics*, 72(1-2):85–134. 46, 72
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–84. 15
- Hamilton, J. D. (2006). Computing power and the power of econometrics. *Medium Econometrische Toepassingen*, 14(2):32–38. 14

REFERENCES

- Hammersley, J. M. and Handscomb, D. C. (1964). *Monte Carlo Methods*. Methuen. 16, 26, 40
- Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, 61(2):247–64. 15
- Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press. 105
- Heckman, J. (1981). Statistical models for discrete panel data. In Griliches, Z. and Intriligator, M. D., editors, *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press. 20, 61
- Heer, B. and Maussner, A. (2005). *Dynamic General Equilibrium Modelling*. Springer, Berlin, Germany. 152
- Hendry, D. F. (1984). Monte carlo experimentation in econometrics. In Griliches, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2 of *Handbook of Econometrics*, chapter 16, pages 937–976. Elsevier. 40, 112
- Himarios, D. (1989). Do devaluations improve the trade balance? the evidence revisited. *Economic Inquiry*, 27(1):143–68. 63, 81, 83
- Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica*, 67(6):1255–1294. 20, 59
- IMF (2003). External debt statistics: guide for compilers and users. Technical report, Washington, D.C. 66
- Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press. 105
- Judd, K. L. (1998). *Numerical Methods in Economics*. The MIT Press. 32, 33, 38, 42, 151, 152, 157, 158
- Judd, K. L. and Guu, S.-M. (1997). Asymptotic methods for aggregate growth models. *Journal of Economic Dynamics and Control*, 21(6):1025–1042. 151

REFERENCES

- Junz, H. B. and Rhomberg, R. R. (1973). Price competitiveness in export trade among industrial countries. *American Economic Review*, 63(2):412–18. 61, 63, 81
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45. 15
- Katos, M. H. and Whitlock, P. A. (2008). *Monte Carlo Methods*. Wiley-Verlag. 35
- Keane, M. P. (1994). A computationally practical simulation estimator for panel data. *Econometrica*, 62(1):95–116. 46
- Kim, H., Kim, J., Schaumburg, E., and Sims, C. A. (2005). Calculating and using second order accurate solutions of discrete time dynamic equilibrium models. Discussion Papers Series, Department of Economics, Tufts University 0505, Department of Economics, Tufts University. 151
- Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *Review of Economic Studies*, 65(3):361–93. 105
- Kitagawa, G. (1987). Non-gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82(400):1032–1041. 15, 22
- Kitagawa, G. (1996). Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25. 105, 109, 110
- Klein, P. (2000). Using the generalized schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics and Control*, 24(10):1405–1423. 153
- Kloek, T. and van Dijk, H. K. (1978). Bayesian estimates of equation system parameters: An application of integration by monte carlo. *Econometrica*, 46(1):1–19. 16, 41

REFERENCES

- Koopman, S. J., Shephard, N., and Creal, D. (2009). Testing the assumptions behind importance sampling. *Journal of Econometrics*, 149(1):2–11. 42
- Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the nelder-mead simplex algorithm in low dimensions. *SIAM Journal on Optimization*, 9:112–147. 129
- Laroque, G. and Salanie, B. (1993). Simulation-based estimation of models with lagged latent variables. *Journal of Applied Econometrics*, 8(S):S119–33. 16
- Lee, L.-F. (1997). Simulated maximum likelihood estimation of dynamic discrete choice statistical models some monte carlo results. *Journal of Econometrics*, 82(1):1–35. 60
- Lehmann, E. and Romano, J. (2005). *Testing Statistical Hypotheses*. Springer. 52
- Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Springer-Verlag New York, Inc., New York, NY. 52, 53
- Lerman, S. R. and Manski, C. F. (1981). On the use of simulated frequencies to approximate choice probabilities. In Manski, C. F. and McFadden, D., editors, *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press. 45
- Liesenfeld, R. and Richard, J.-F. (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance*, 10(4):505–531. 15, 16
- Liesenfeld, R. and Richard, J.-F. (2007). The multinomial multiperiod probit model: Identification and efficient estimation. Technical report. 49, 76, 101
- Liesenfeld, R. and Richard, J.-F. (2008a). Improving mcmc, using efficient importance sampling. *Computational Statistics & Data Analysis*, 53(2):272–288. 18, 19
- Liesenfeld, R. and Richard, J.-F. (2008b). Simulation techniques for panels: Efficient importance sampling. In Matyas, L. and Sevestre, P., editors, *The Econometrics of Panel Data*. Kluwer Academic Publishers. 15, 16, 69

REFERENCES

- McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5):995–1026. 15
- McGrattan, E. R. (2006). Notes: applying the finite element methods in dynamic economies. Technical report, Federal Reserve Bank of Minneapolis. 157
- Mendoza, E. G. (1991). Real business cycles in a small open economy. *American Economic Review*, 81(4):797–818. 23, 118, 138, 142, 167
- Metropolis, N. (1987). The beginning of the monte carlo method. *Los Alamos Science*, pages 125–130. 35, 36
- Metropolis, N. and Ulam, S. (1949). The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341. 15, 36
- Milesi-Ferretti, G. M. and Razin, A. (1996). Current account sustainability. Princeton Studies in International Finance 81, Princeton University. 62
- Milesi-Ferretti, G. M. and Razin, A. (1997). Current account sustainability: Selected east asian and latin american experiences. NBER Working Papers 5791, National Bureau of Economic Research, Inc. 60
- Milesi-Ferretti, G. M. and Razin, A. (1998). Sharp reductions in current account deficits an empirical analysis. *European Economic Review*, 42(3-5):897–908. 60, 63, 65, 66, 67, 80
- Milesi-Ferretti, G. M. and Razin, A. (2000). Current account reversals and currency crises, empirical regularities. In *Currency Crises*, NBER Chapters, pages 285–326. National Bureau of Economic Research, Inc. 60
- Mirana, M. J. and Fackler, P. L. (2002). *Applied Computational Economics and Finance*. The MIT Press. 32, 33, 151, 152, 157
- Obstfeld, M. and Rogoff, K. (2007). The unsustainable u.s. current account position revisited. In *G7 Current Account Imbalances: Sustainability and Adjustment*, NBER Chapters, pages 339–376. National Bureau of Economic Research, Inc. 60

REFERENCES

- Obstfeld, M. and Rogoff, K. S. (1996). *Foundations of International Macroeconomics*, volume 1 of *MIT Press Books*. The MIT Press. 68
- Pitt, M. K. (2002). Smooth particle filters for likelihood evaluation and maximisation. Technical report. 106, 110, 112
- Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599. 106, 110
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes. The Art of Scientific Computing*. Cambridge University Press, third edition edition. 29, 31, 32, 33, 39, 157
- Richard, J.-F. (1977). Bayesian analysis of the regression model when the disturbances are generated by an autoregressive process. In Aykac, A. and Brumat, C., editors, *New developments in the application of Bayesian methods*. North Holland, Amsterdam. 77
- Richard, J.-F. and Zhang, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141(2):1385–1411. 18, 42, 44, 49, 51, 52, 55, 60, 74, 77, 106, 113, 114, 136, 165
- Ristic, B., Arulampalam, S., and Gordon, N. (2004). *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House. 106
- Robert, C. P. and Casella, G. (2005). *Monte Carlo Statistical Methods*. Springer. 17, 26, 41
- Sargent, T. J. (1989). Two models of measurements and the investment accelerator. *The Journal of Political Economy*, 97(2):251–287. 22, 107, 148
- Schmitt-Grohe, S. and Uribe, M. (2003). Closing small open economy models. *Journal of International Economics*, 61(1):163–185. 23, 118, 138, 142, 167
- Schmitt-Grohe, S. and Uribe, M. (2004). Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control*, 28(4):755–775. 139, 149, 151, 153

REFERENCES

- Shephard, N. (1993). Fitting non-linear time series models, with applications to stochastic variance models. *Journal of Applied Econometrics*, 8:S135–52. 16
- Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84:653–667. 49
- Sims, C. A. (2002). Solving linear rational expectations models. *Computational Economics*, 20(1-2):1–20. 153
- Smets, F. and Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, 1(5):1123–1175. 107, 138
- Smith, J. and Santos, A. A. (2006). Second-order filter distribution approximations for financial time series with extreme outliers. *Journal of Business & Economic Statistics*, 24:329–337. 112
- Stern, S. (1997). Simulation-based estimation. *Journal of Economic Literature*, 35(4):2006–2039. 15, 59
- Stroud, A. H. (1971). *Approximate Calculation of Multiple Integrals*. Prentice-Hall. 19, 26, 34
- Stroud, A. H. and Secrest, D. (1966). *Gaussian Quadrature Formulas*. Prentice-Hall. 26, 31, 32, 33
- Taylor, J. B. and Uhlig, H. (1990). Solving nonlinear stochastic growth models: A comparison of alternative solution methods. *Journal of Business & Economic Statistics*, 8(1):1–17. 23
- Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press. 46
- Uzawa, H. (1968). Monte carlo experimentation in econometrics. In Wolfe, J., editor, *Value, Capital and Growth: Papers in Honor of Sir John Hicks*, pages 485–504. The University of Edinburgh Press. 138
- van Dijk, H. K. (1999). Some remarks on the simulation revolution in bayesian econometric inference. *Econometric Reviews*, 18(1):105–112. 16, 17, 18, 49, 165

REFERENCES

- Whittle, P. (1983). *Prediction and Regulation by Linear Least-Square Methods*. Basil Blackwell Publisher. 39
- Winkelmann, R. and Boes, S. (2006). *Analysis of microdata*. Springer. 93
- Zhang, W. and Lee, L. (2004). Simulation estimation of dynamic discrete choice panel models with accelerated importance samplers. *Econometrics Journal*, 7(1):120–142. 76

Guilherme Valle Moura

Vrije Universiteit Amsterdam
Department of Econometrics & OR
De Boelelaan 1105
1081 HV Amsterdam
Room 1A-30

gvallemoura@gmail.com
gmoura@feweb.vu.nl

Personal Details

- **Date of birth:** January 2nd, 1981.
- **Citizenship:** Brazilian.
- **Gender:** Male.
- **Marital Status:** Married, no children.

Education

- **University of Kiel**, Germany.
PhD in Economics (*Dr. rer. pol.*), October 2005 - January 2010.
Dissertation title: Efficient Importance Sampling in Applied Econometrics.
- **University of Pittsburgh**, USA. Visiting Scholar, February 2008.
- **Federal University of Rio Grande do Sul**, Brazil.
M.Sc. in Economics, March 2003 - March 2005.
- **Pontifical Catholic University of Minas Gerais**, Brazil.
B.Sc. in Economics, February 1999 - December 2002.
- **Scholarships**
 - DAAD Jahresstipendium (April 2006 - December 2010)
 - University of Kiel Scholarship (October 2005 - March 2006)
 - CAPES-Brazil Scholarship (April 2003 - March 2005)
 - PROBIC Scientific Initiation Scholarship (February 2002 - December 2002)

Academic Positions

- **Assistant Professor**, Vrije Universiteit Amsterdam (January 2010 -)
- **Research Assistant**, University of Kiel (July 2006 - June 2009)
Project: "Panel Data Models with Generalized Error Structures for the Analysis of Current Account". Funded by the Deutsche Forschungsgemeinschaft.
- **Teaching Assistant**, University of Kiel (Winter semester 2008-2009)
Course: Statistics of Financial Markets.
- **Lecturer**, Pontifical Catholic University of Minas Gerais (February 2005 - July 2005)
Courses: Applied Economics, Introduction to Economics, and Research Methodology.
- **Teaching Assistant**, Federal University of Rio Grande do Sul (February 2004 - July 2004)
Course: International Economics I.

Research

- **Publications**
 - Determinants and Dynamics of Current Account Reversals: An Empirical Analysis, with Roman Liesenfeld and Jean-François Richard. Oxford Bulletin of Economics and Statistics, *forthcoming*.

- Does the Expiration of Dollar Indexed Bonds Affect the Exchange Rate? An Econometric Event Study, written in portuguese with Roberto Meurer and Maurício Nunes. *Revista de Economia Aplicada*, v. 11, p. 55-72, 2007.
- Travel Hysteresis in the US Current Account After the Mid-1980s, with Sergio Da Silva and Roberto Meurer. *Economics Bulletin* v.14, n.2, p. 1-10, 2005.
- Travel Hysteresis in the Brazilian Current Account, with Sergio Da Silva and Roberto Meurer. *Economics Bulletin* v.6, n.24, p. 1-17, 2005.
- Big Mac Parity, Income, and Trade, with Sergio Da Silva and Sidney Caetano. *Economics Bulletin*, v.6, n. 11, p. 1-8, 2004.
- Is There a Brazilian J-Curve? With Sergio Da Silva. *Economics Bulletin*, v. 6, n. 10, p. 11-17, 2005.

· **Working Papers**

- Efficient Likelihood Evaluation of State-Space Representations, with David DeJong, Hariharan Dharmarajan, Roman Liesenfeld and Jean-François Richard. (Revised and resubmitted to *The Review of Economics Studies*)
<http://www.stat-econ.uni-kiel.de/team/EWP-2009-02.pdf>

· **Interests**

- Simulation Based Inference
- Nonlinear State Space Models
- Nonlinear Panel Data Models

Conferences

- **Statistische Woche**, Meeting of the German Statistical Society. Dynamic Panel Probit Models for Current account Reversals and their Efficient Estimation, 24-27 September 2007, Kiel, Germany.
- **CFE08**, 2nd International Conference on Computational and Financial Econometrics. Dynamic Panel Probit Models for Current account Reversals and their Efficient Estimation, June 2008, Neuchâtel, Switzerland.
- **LAMES**, Latin American Meeting of the Econometric Society. Efficient Likelihood Evaluation of Nonlinear RBC Models, November 2008, Rio de Janeiro, Brazil.
- **ESEM**, European Meeting of the Econometric Society. Efficient Likelihood Evaluation of State-Space Representations, August 2009, Barcelona, Spain.
- **CFE09**, 3rd International Conference on Computational and Financial Econometrics. Efficient Likelihood Evaluation of State-Space Representations, October 2009, Limassol, Cyprus.

Languages

- **Portuguese**: Native speaker.
- **English**: Fluent. Certificate in Advanced English from the University of Cambridge.
- **German**: Fluent. Basic, intermediate and advanced courses at the University of Kiel.
- **Spanish**: Fair. Basic course.

Other skills

- MATLAB, C/C++, Stata, Eviews, VBA, L^AT_EX, Microsoft Office.

References

- **Professor Roman Liesenfeld** (Advisor)
Institute for Statistics and Econometrics
University of Kiel
Olshausenstraße, 40-60
D-24118, Kiel - Germany
Phone:(+49) 431-880-2166
E-mail: `liesenfeld@stat-econ.uni-kiel.de`

- **Professor Jean-François Richard**
Department of Economics
University of Pittsburgh
4711 Wesley W. Posvar Hall
Pittsburgh, PA 15260 - USA
Phone:(+1) 412-648-1750
E-mail: `fantin@pitt.edu`

- **Professor David N. DeJong**
Department of Economics
University of Pittsburgh
4711 Wesley W. Posvar Hall
Pittsburgh, PA 15260 - USA
Phone:(+1) 412-648-2242
E-mail: `dejong@pitt.edu`