The Empirical Analysis of Exchange Rate Regimes and Nonlinear Econometrics

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

Viktor Winschel

vorgelegt im Wintersemester 2005

Referent:	Prof. Dr. Roland Vaubel
Korreferent:	Prof. Felix Kübler, Ph.D.
Dekan:	Prof. Konrad Stahl, Ph.D.
Tag der mündlichen Prüfung:	19.12.2005

Acknowledgements

I thank Prof. Dr. Roland Vaubel for his patience during this Ph.D. research and for the insights into economics and economic policy. I also profited from Wouter Denhaan and Michel Juillard and their seminars on computational economics taught at the University of Mannheim. Kenneth Judd and James Heckman gave me interesting hints at the Institute on Computational Economics at the Argonne National Laboratory in Chicago. Florian Heiss was an invaluable discussant in never-ending debates on econometrics, economics and the rest. I also want to thank my parents for their great support during all the years. And last, but not least, I thank my wife who always encouraged me when I was in trouble.

Viktor Winschel

.

The dissertation includes the following four papers:

EMU - Enlargement

SVAR - Decomposition:
Does the Nominal Exchange Rate Facilitate Real Exchange Rate Adjustment in Floating Exchange Rate Regimes?
A SVAR Decomposition of German Exchange Rates in 1973-1998

Macroeconometrics:
Solving, Estimating and Selecting Nonlinear Dynamic Economic Models without the Curse of Dimensionality

Microeconometrics:
Smolyak Cubature for Multiple Integration in Estimation Problems

.

Contents

.

Chapter 1

Introduction

In the first paper in chapter 2 (EMU-Enlargement), I use the traditional optimum currency area theory to derive a suitable group of countries for the enlargement of the European monetary union. This informal theory does not provide an absolute quantitative measure for the welfare gains of a monetary union. The methodological conclusion is that a microfounded, structural approach is indispensable for a conclusive policy recommendation. Moreover it is a way to cope with the endogeneity problem inherent in this field.
In the second paper in chapter 3 (SVAR decomposition), I ask whether flexible nominal exchange rates facilitate the real adjustment. For this purpose I decompose the bilateral German nominal exchange rates. The driving demand, supply and monetary shocks are identified by a long-run restriction scheme within a SVAR approach. This paper is half the way towards a structural analysis of the real adjustment mechanism underlying the choice of an optimal currency area or exchange rate regime.
The need to evaluate welfare effects of the real adjustment and the exchange rate policy led to the main effort of this dissertation in chapter 4 (Macroeconometrics). In this paper I develop a Bayesian estimation framework for nonlinear models suitable for a quantitative empirical macroeconomic welfare analysis. The mathematical and computational challenges were considerable and so far I have not used the software for a structural analysis of the real adjustment performance of alternative exchange rate regimes.
The last paper of this dissertation (Microeconometrics) is an offspring of my studies of the Smolyak operator for the nonlinear model estimation. It is a joint paper with Florian Heiss where we applied the operator for the estimation of a microeconometric mixed logit model. Florian contributed his microeconometric expertise and me the Smolyak operator. The mathematical problem is a high dimensional integral defining the likelihood of the model. We show in extensive Monte Carlo simulations that parameter estimates based on a Monte Carlo integration are dramatically outperformed compared to an estimation based on Smolyak Gaussian quadrature.

Chapter 2

EMU - Enlargement

2.1 Introduction

The suitability of countries for an exchange rate system can be analyzed from two different perspectives. The first is to derive the optimal exchange rate system for a given group of countries. This is not the strategy of this paper.
The other perspective is to identify suitable countries for a common currency according to the theory of optimal currency areas (OCA).[1] This is an economic cost-benefit analysis of adopting a monetary union. Political considerations are not taken into account. For example membership in the European Union (EU) and the European Monetary Union (EMU) could stabilize the transformation process in the former socialist countries. An opposite political deliberation could be that a loss of monetary sovereignty is opposed.[2] This may be even more important in the former socialist countries which attained sovereignty only recently. The theory of optimal currency areas is Paretian in the sense that the question is not which particular country loses or wins, but which group of countries as a whole loses or wins when forming a single currency area. The theory was developed in the sixties and discusses the costs of macroeconomic management unification. Critics argue that the costs of a unification are lower than traditionally suggested[3]- since the criteria for joining a monetary union are endogenous[4]. The focus of this paper is on the application of the empirical criteria for a cost-benefit analysis of European countries, which are classified in four groups: **EU11**: members of the EMU: Austria, Finland, France, Germany, Ireland, Italy, Luxem-

[1] The theory of OCA was introduced by Mundell (1961), McKinnon (1963), Kenen (1969).

[2] Like a Dane, who was interviewed during the Danish Euro referendum in Brown-Humes and MacCarthy (2000): '"It's not just about the single currency. Little by little, we are being eaten up by Europe," he says. The fear that joining the Euro will make Danes lose ever more of their national identity helps to explain why the outcome of the today's Euro referendum is on a knife-edge.'

[3] Grauwe (1997), Chapter 3, p. 20-50

[4] Frankel and Rose (1998)

bourg, Belgium, the Netherlands, Portugal, Spain, **EU+4**: members of the EU, but not of the EMU: Denmark, Greece[5], Sweden, the United Kingdom, **CEE5**: countries with accession talks for the EU: the Czech Republic, Estonia, Hungary, Slovenia, Poland, **CEE+8**: other candidates for the EU: Cyprus, Malta, Turkey, Romania, Bulgaria, the Slovak Republic, Lithuania and Latvia.
The data base spans over the period from the beginning of the nineties until the end of 1998, just before the start of the Euro in January 1999. The first criterion is the historical record of real exchange rate changes. It represents a comprehensive measure of the causes of real exchange rate changes; namely trade specialization, shock correlation and economic policy shocks. This enumeration cannot be exhaustive. The emphasis lies on historical real exchange rate changes since this criterion includes *and* weights the enumerated as well as all other causes. Thereafter, those adjustment instruments are examined which might be able to compensate for lack of nominal exchange rate adjustments; namely the factor mobility and fiscal transfers. Their functioning can also be measured by the real exchange rate criterion. Subsequently the cost factors are criticized, benefits are discussed and finally costs and benefits are weighted in order to arrive at the conclusion about the qualification of each particular country for the EMU.

2.2 Costs of a monetary union

2.2.1 Historical real exchange rate changes

The abandonment of nominal exchange rate changes implies costs which can be approximated by the historical record of real exchange rate changes.[6] The relative price between commodity baskets has to adjust to demand and supply changes. In a system of flexible exchange rates this is done by changes in the nominal exchange rate. The advantages of such a system are threefold: instantaneous adjustment, internal price stability is possible despite relative price changes and relative wages are adjusted likewise through the nominal exchange rate. Since nominal exchange rates cannot adjust in a monetary union, the need for real exchange rate changes has to be satisfied by differences in price changes, like inflation rates and wage changes.

[5] Even though Greece has been a member of the EMU since 1.1.2001 it will be classified into the EU+4 group to test for the suitability of this enlargement, as this analysis was completed in 2000.

[6] This method has been introduced by Vaubel (1976).

The historical need for real exchange rate changes is approximated by the standard deviation of real exchange rate changes according to the following formula:

$$\sigma_R = \sqrt{\frac{\sum_{i=1}^{n}(\pi_i - \bar{\pi})^2}{n-1}} \tag{2.1}$$

$$\pi_i = \sqrt[l]{\frac{E_{i,t}P_{i,t}}{E_{i,t-l}P_{i,t-l}}} - 1 \tag{2.2}$$

$$\bar{\pi} = \frac{1}{n}\sum_{i=1}^{n}\pi_i,$$

where E_i is the nominal exchange rate of country i, measured in Dollar per national currency unit, P_i is the consumer price level in country i and n is the number of countries in the group.
To be more precise: the π_is are inflation rates measured in a common currency. Since all real exchange rates are measured in relation to the US, its price level P^* in the usual real exchange rate definition can be dropped because only the relative standard deviation of the cross sectional real exchange rate changes matters and the order in differently sized groups (different n) is not affected by this scaling parameter.
The standard deviation $\sigma_{R,USA}$ in the US is used as the benchmark of an existing monetary union. The π_is are replaced by the US regional inflation rates.[7] Since the P_is are already measured in the same currency, multiplication by a numeraire exchange rate is not necessary.
An important decision concerns the length of the period for which the growth rates are calculated. For too short periods, real exchange rates may contain monetary shocks, since nominal exchange rates react faster than prices. This influence of the prevailing monetary system should not be allowed to affect the suitability for another monetary system. For periods too long, the variability of the causes of real exchange rate changes will be understated.
In order to take this problem into account two different approaches are conceivable. The first approach, as used in Vaubel (1976), chooses the period length to permit prices to adjust. For fitted values of the π_is, the real exchange rate movements due to different reaction lags are hopefully negligible. The subsequently used π_is have been fitted on a yearly basis in order to keep the results comparable among periods with different length.
In this study an additional calculation will be done. A structural vector autoregression (SVAR) is used to decompose the real exchange rate into monetary and real shocks by Blanchard and Quah (1989) type long-run restrictions.[8] The idea is to subtract monetary effects from the real exchange rate and thereby to allow

[7] See appendix 2.6.3 for the regions used for this purpose.
[8] Applied for instance by Enders and Lee (1997) or Dibooğlu and Kutan (2000).

for the endogeneity of the monetary component. The decomposition shows that monetary effects account only for a minor part of real exchange rate movements so that total real exchange rate changes may be used as well. This is a rough first approach, which entails the basic features for a more elaborated one with structural models and economically derived identifying restrictions.
The estimated SVAR can be represented in a matrix notation as

$$\begin{aligned} \mathbf{x}_t &= \mathbf{B}_1\mathbf{x}_{t-1} + \mathbf{B}_2\mathbf{x}_{t-2} + ... + \mathbf{B}_n\mathbf{x}_{t-n} + \mathbf{e}_t \\ &= [\mathbf{I} + \mathbf{B}(L) + \mathbf{B}(L)^2 + ...]\mathbf{e}_t, \end{aligned} \tag{2.3}$$

where $\mathbf{x}_t$ is the vector comprising Δr_t and Δp_t, the changes of natural logarithms of the real exchange rate and the consumer price index. The real exchange rate was calculated using the Dollar nominal exchange rate and consumer price indices. $\mathbf{B}_i$ are the estimated coefficient matrices, $\mathbf{e}_t$ the residuals, and $\mathbf{B}(L)$ is a polynomial in the lag operator. The residuals in (2.4) are independent nominal and real shocks $\boldsymbol{\varepsilon}_t = [\varepsilon_{nt}\ \varepsilon_{rt}]'$ and thus the estimated reduced form residuals have to be orthogonalized by $\mathbf{e}_t = \mathbf{C}\,\boldsymbol{\varepsilon}_t$ to structural shocks. For a two dimensional $\mathbf{x}_t$ four restrictions are needed to identify $\mathbf{C}$ and to arrive at the structural vector moving average representation:

$$\begin{bmatrix} \Delta r_t \\ \Delta p_t \end{bmatrix} = \sum_{i=0}^{\infty} L^i \begin{bmatrix} a_{11i} & a_{12i} \\ a_{21i} & a_{22i} \end{bmatrix} \begin{bmatrix} \varepsilon_{nt} \\ \varepsilon_{rt} \end{bmatrix}. \tag{2.4}$$

Two out of four identifying restrictions are normalizations defining the variance of the shocks ε_{dt} and ε_{st}. The third assumes orthogonality and the fourth

$$\sum_{i=0}^{\infty} a_{11\,i} = 0 \tag{2.5}$$

represents the assumption that nominal shocks have no permanent effects on the real exchange rate. The results of this decomposition are presented in figures in section 2.6.1. The figures show the actual and the decomposed real exchange rates for each country. The decomposed rates are driven only by real shocks or nominal shocks. The result is that the real exchange rate is determined almost exclusively by real determinants. It is seen as evidence in favor of the method of Vaubel (1976), where, like here in the following, the total real exchange rate is used.
In interpreting the results, one should take into account that a monetary union may positively affect the factor mobility, trade diversification, the openness of countries and the business cycle correlation so that the optimum currency area criteria are endogenous. The objection could be tested for Europe by comparing these variables in the pre- and post-European monetary unification periods. But the post-unification period is still too short to do this. Therefore, more conclusive direct empirical results cannot yet be derived. However, the second half of

the endogeneity argument can be tested, namely that trade integration diversifies trade flows. The first half is discussed in Rose (2000), who evaluated the effect of a common currency on trade for 180 countries. The results show a huge positive effect. He also tested, whether trade between countries with a common currency is redirected at the expense of trade with other countries. This hypotheses is rejected. The full argument runs as follows: a common currency diversifies the trade flows, reduces the likelihood of asymmetric shocks and therefore the need for real exchange rate changes. Thus a common currency makes a country more suited for a common currency so that the optimal currency area criteria are endogenous. A quantitative answer to this objection can only be given by an analysis with structural models and counterfactual simulations. A simpler question whether trade integration diversifies the trade flows will be addressed in section 2.2.4.1.

The fitted real exchange rate changes for each country are reported in Table 2.1, p. 27. The calculated figures show strongly divergent values for the CEE and EU countries (except for Turkey, Malta, Cyprus and Hungary), as can be seen in the lower part of Table 2.1. Therefore one can suspect that are less suitable members of a common currency area.

In Table 2.2, p. 28 these standard deviations are calculated for different groups and time periods. In order to facilitate the comparison, they are reported as multiples of a benchmark standard deviation (USA11 in 1992). They tend to grow from the EU11, EU11+4 to CEE5 and CEE+8. The standard deviations of USA11 (lower part, the seven one-year-periods) remain between 0.6 and 1.2 for all periods. A smoothing effect can be detected for the EU11 countries, but not for the CEE countries, where the standard deviations of the one year periods are not monotonically decreasing.

In transition economies sharp inflation increases can be ascribed to price liberalization and the destruction of an inherited monetary overhang because of forced savings. The huge standard deviations in 1992 and 1993 for the CEE countries are due to the Baltic hyperinflation of around 1000%, which was not offset by the nominal exchange rate depreciation. Therefore the real exchange rate changed substantially. The rising consumer prices in Estonia can be traced back to the increased regulated prices in protected sectors.[9] During this time the Baltic currencies were established and Lithuania's convertibility restrictions were abolished. In Latvia, a crisis of the banking system took place in 1995. After the collapse of the biggest commercial bank 35% of deposits were withdrawn from solvent banks. Similar events took place in Lithuania.

The large standard deviations in 1997 are caused by Romania and Bulgaria. The lax monetary policy since 1995 resulted in a crisis in 1997. In Bulgaria and Romania state loans to weak banks and companies increased the money stock.

[9] Hasselblatt, Proos, and Zirnask (1996), p.63

Table 2.3, p. 29 gives an idea of the burdens of relative price changes in each country based on the seven-years-period assuming a zero inflation in different monetary unions and long run monetary neutrality on real exchange rates.
So far the countries were grouped indiscriminately. In Table 2.4, p. 30, the least suitable countries are successively excluded so that the remaining countries exhibit the smallest possible real exchange rate standard deviation. The costs are thus measured by the relative price movements adjusting fully to the need of real exchange rate changes when nominal exchange rates are fixed. The USA real exchange rate changes are taken as a benchmark for a group of countries to be classified as an optimal currency area.
I shall explain the notation if Table 2.4, p. 30 with reference to the first columns. 10.56 is the standard deviation of all countries. Excluding Lithuania results in a standard deviation of 8.12. Ireland and Luxembourg is the pair of countries giving the smallest possible standard deviation. The lower part of the table compares the $\sigma_{R,i}$s with $\sigma_{R,USA} = 0.41$. Calculating the F-test $F_{USA/Ita} = \frac{0.41^2}{0.26^2} \approx 2.457$ asks whether $\sigma_{R,i}$ of the group without Italy (0.26) is significantly lower than the one of the US regions (H_0: equality, $H_1 : \sigma^2_{R,USA} > \sigma^2_{R,i}$). The p-value for this test is 7.8%=0.078. Therefore Malta,..., Luxembourg are a less desirable currency area than the U.S. Excluding Malta results in a F-Test of 4.072 with a p-value of 1.8%=0.018. This remaining currency area including Denmark,..., Luxembourg is highlighted by bold typed country names. Countries written in italic are significant at the 10% level.
The columns denoted by "All" successively exclude countries until two are left. This answers the question for a suitable monetary union regardless of the existing EMU. The columns denoted by "EU11+" exclude countries given that the EMU is already implemented for the EU11 countries. The bold typed countries if included would not increase the standard deviation beyond the EU11 level and are therefore taken as possible candidates for the enlargement: Denmark, Sweden and Greece. The results for Turkey's demonstrate the effect of the choice of a certain period length to calculate the real exchange rate changes: its real exchange rate increased substantially in the first 4 month of 1994 and decreased until the end of 1995 to the end-of-1993 value and again exhibits a substantial change by the end of 1998. As a consequence, between 1992 and 1998, the real exchange rate change is smoothed. In the 1994-1998 period the measured change captures the recovery from the end-of-1993 change and 1996-1998 is dominated by the end of 1998 change. Spain exhibits a similarly volatile real exchange rate. It disqualifies Spain after the UK in the 1992-1998 column, but not in the other two periods.
The main result does not significantly depend on the chosen period. The Baltic countries are the least suitable candidates, but they get more suitable, the fewer years are taken into account from the early decade. In the period 1996-1998 Bulgaria and Romania are disqualified very fast, due to their lax monetary policy in 1997 and 1998. The core western Europe of Luxembourg, Belgium, Austria and Germany are always among the most suitable countries. Here the endogeneity

objection comes from the formation of a DM-zone. The least suitable country in the EU11+4 group is the United Kingdom. Malta, Cyprus, Slovenia, and Hungary are the most suitable CEE countries for the EMU membership.
It is interesting to see that the order of countries hardly changes comparing the 17 top rows of the columns All and EU11+ for each period. It has to be compared from the top, since the order of the EU11 countries in the EU11+ columns is arbitrary. The two orders are the same until the first horizontal line from the top. The second line delimits them from the EU11. Countries listed in only one of the two columns between these two lines are set one blank to the right indicating the effect of the EMU11 restriction. In the All columns these countries are Spain, Italy and Ireland meaning that they would be excluded from an eleven country monetary union by the real exchange rate criterium, if members were not restricted to the starting EU11 group. In the EU11+ columns the countries Denmark, Turkey, Sweden and Greece are suitable for an eleven country union. They are not only in the variance minimizing eleven countries group (indicated by a right shift) but also, except for Denmark, in the left seven-years-period, bold faced. Meaning that they would not enlarge the standard deviation of real exchange rate changes in the EU11 monetary union. However, this result depends on the chosen period and thus on the smoothing effect of the period length.

2.2.2 Causes of real exchange rate changes

2.2.2.1 Trade specialization

Low trade specialization or high diversification may imply shock compensation and therefore a small need for a real exchange rate adjustment. Low specialization is an indicator of a high monetary union suitability. It is not the sectoral specialization, but the specialization of import and export flows which is relevant to this argument.
Disaggregated data on commodity flows and a trade partner are available in the expensive United Nations COMTRADE database. A less suitable database is the Eurostat's COMEXT CD-ROM which was used in earlier calculations and allows to specify trade flows from the EU countries to other countries. This database neglects trade flows among non-EU countries.
The data set used here is a freely available part of the COMTRADE database disaggregated into roughly 260 SITC (Standard International Trade Classification) commodity categories but not by trade partners. To assess the bias resulting from trade flows with the rest of the world instead of restricting them to the EU and CEE countries, I compared the results from this aggregated COMTRADE and the COMEXT database for the year 1995. The differences are not serious and the aggregated COMTRADE 1998 figures are taken as a reasonable approximation to the COMEXT and the disaggregated COMTRADE database.

The trade flow specialization is measured by:

$$\text{SI} = \frac{n}{2(n-1)} \sum_{i=1}^{n} \left| a_i - \frac{1}{n} \right|, \tag{2.6}$$

where a_i is the share of a commodity category in export or import and n is the number of categories. The idea is to calculate the deviation from a uniformly distributed trade flow. Perfect diversification results in SI=0 since each category has a share of $\frac{1}{n}$. Perfect specialization in only one category results in a SI of 1. An implicit assumption is that trade shares in a commodity category are uniformly distributed. Therefore the order of the countries is not invariant to the number of categories used. The aggregation of the export and import specialization to trade specialization, the variable of interest, is done by simple averaging. Weighting according to import and export shares does not dramatically change the result. Another possibility is to calculate the specialization of net exports.
Figure 2.1, p. 31 plots both specialization indices, which are averaged in Table 2.5, p. 31. The peripheral countries show a tendency to specialization. If only trade flows among the EU and CEE countries are used, specialization may have been explained by smaller trade flows of peripheral countries.
Cyprus, Ireland and Malta are the most endangered countries with respect to shocks. The Czech Republic, France, Belgium/Luxembourg,... have the most diversified trade flows. Spain as a rather peripheral country has a low trade specialization. Poland, Slovenia, Slovakia and Estonia from the group of non EMU members are most diversified.
Capital flows also imply real exchange rate adjustments. From the identity of the current and capital account balance X-Im+Tr $\equiv \Delta$NEP it is clear that the real exchange rate is affected by both sides of the account. A growing capital account balance implies a growing current account balance and real devaluation.
Capital imports are most important in the former socialist countries to assist the transformation purposes. Table 2.6, p. 32 presents the foreign direct investments (FDI) in Europe. These flows might strongly depend on viable capital markets and banking sectors, the progress of privatization or general development. But there might be a simultaneous relation. Indices of the transition progress are shown in Table 2.15, p. 41.
Since non of the transition countries finished the reform process yet, future FDIs and thus a high need for real exchange rate adjustment can be expected.

2.2.2.2 Shock correlation

The shock correlation is an important criterion since a monetary union abandons country specific monetary policies.
Supply and demand shocks have to be distinguished, since a high correlation of demand shocks in one country with supply shocks in another country implies the

need for high real exchange rate adjustments, whereas a high correlation of two likewise shocks implies only a small need for adjustment.
Here I use a method also applied by Bayoumi and Eichengreen (1994) to an optimal currency area discussion. It is again the structural VAR approach developed by Blanchard and Quah (1989) and adopted to the needs of the OCA theory by Bayoumi (1992) to identify aggregate supply and demand disturbances.
It is difficult to arrive at an appropriate solution for the starting year of these calculations. Before 1990 the socialist countries had a different economic structure and the data is not likely to be representative for the future. The problem with the data after 1990 is that the very purpose of a transformation process is to change the economic structure and that measured correlations may reflect this process rather than the structure. This objection is valid for any post transition year until the full catch up. As a compromise the period 1990-1998 was used.
The estimated SVAR is again given in equation (2.3) with different variables: vector $\mathbf{x}_t$ includes Δy_t and Δp_t, the natural logs of the industry production index and the producer price index. The residuals in (2.3) are orthogonalized by $\mathbf{e}_t = \mathbf{C}\,\boldsymbol{\varepsilon}_t$ in order to identify demand and supply shocks $\boldsymbol{\varepsilon}_t = [\varepsilon_{dt}\ \varepsilon_{st}]'$.
Again three out of four identifying restrictions are the variance normalization and orthogonality assumption. The fourth restriction is

$$\sum_{i=0}^{\infty} a_{11i} = 0, \tag{2.7}$$

and it assumes that demand shocks have no permanent effects on output.
For the period 1990-1998 GDP figures would provide only a few observations, since quarterly data is not available for all countries and all years. Therefore industrial production indices were used even though this accounts only for a fraction of GDP. Table 2.7, p. 33 shows the sectoral shares of the agricultural, industrial and service activities, sorted by the sectoral specialization index SI, as defined in equation (2.6). The production indices for Estonia and Malta were not available.
Like in Bayoumi and Eichengreen (1994) the significance was measured by an assumed null value of the correlation in order to "exclude that part accounted for by the international business cycle, for only deviations from common movements are important in assessing the suitability of a group of countries for monetary unification."[10] The correlation $\rho = 0.05$ between Germany and Japan was chosen as the null correlation for supply and demand shocks.
The results are shown in Table 2.8, p. 34 and Table 2.9, p. 35. The possible groups are highlighted by frames around submatrices. Three different groups emerge for the supply shocks in Table 2.8. They overlap for some countries but

[10] The test statistic is $0.5\ln[(1+r)/(1-r)] \sim \mathrm{N}\left(0.5\ln[(1+\rho)/(1-\rho)];\sqrt{1/(T-3)}\right)$, where r is the measured and ρ is the null value of the correlation coefficient.

are quite clearly distinguished, since the lower left submatrix and the two bottom lines show almost exclusively negative correlations. The core European group includes France, Belgium, Denmark, Germany, the Netherlands and Austria and was placed in the middle of the correlation matrix. Possible candidates were placed to the upper left and lower right. The upper left group shows some south European countries, the lower right part some north European, but these two groups are not clearly separable by neighborhood. Turkey and Luxembourg are negatively correlated with almost all countries. Classifying the countries in case of supply shocks results in three groups suitable for different monetary unions. All countries except Portugal, Latvia, the Slovak Republic, Lithuania, Luxembourg and Turkey are classified into the unions: the core (C) one, and two more enclosed in the inner frames in the upper left (L) and lower right (R) corners.
The grouping for demand shocks in Table 2.9 is very clear. The shocks of the core European group are again positively correlated even though Belgium shows a minor negative correlation with Austria. The first countries to include are Spain and the UK. Then comes the south group: Greece, Portugal, Turkey and Cyprus and the northern countries Sweden and Finland. The next enlargement includes the countries with a clear negative correlation to the countries so far: Poland, the Slovak Republic, Lithuania, Ireland, Slovenia. Luxembourg again shows strong negative correlations with almost all countries besides the Netherlands, Belgium and some others. The group of Latvia, Bulgaria, Italy, Hungary, Romania, and the Czech Republic is almost completely negatively correlated with all other countries but shows some positive correlation among each other. The countries are classified in a core (C), an extended core (E, in the inner frame of the upper left corner) and a peripheral (P, in the lower right corner) group.
The overall groups from supply and demand shock correlations can be found for countries with the same classification for both shocks. There are four groups emerging for countries which are jointly suitable in both kind of shock correlations: (CC)-Austria, Belgium, France, Germany, the Netherlands, Denmark, (LE)-Spain, Greece, he United Kingdom, Cyprus, (RE)-Finland and Sweden and (RP)-Italy, Bulgaria, Hungary, Slovenia, Romania. Only the Czech Republic with grouping (LP) cannot be unionized. The groups seem to be strongly defined by neighborhood except for the United Kingdom in (LE).
Some causes for shock correlations are discussed and evaluated by simple regressions on bilateral figures. A high supply shock correlation can be explained by per capita GDP. A high income indicates a similar economic structure: a significant (t-statistics in parenthesis) positive relation is found between the specialization index of sectoral shares (SI_{sector}) of GDP and GDP_{PPP} per capita:

$$\mathrm{GDP}_{PPP}\ \text{p.c.} = \underset{(-2.5)}{-14393.7} + \underset{(5.2)}{62482.0}\ \mathrm{SI}_{sector}$$

This reflects the fact that high developed countries have a high share of services in GDP. However, differences in economic structure (coefficient of variation of

sectoral specialization indices) do not have a significantly negative effect on the correlation of supply shocks (ρ_{supply}):

$$\rho_{supply} = \underset{(5.45)}{0.08} - \underset{(-1.24)}{0.09} \text{ difference in economic structure.}$$

The above identified groups suggest that the supply shock correlation can be related to neighborhood, here measured by the distance between main industrial regions:

$$\rho_{supply} = \underset{(5.03)}{0.1} - \underset{(-2.18)}{0.03} \text{ distance in '000 km.}$$

Frankel and Rose (1999), who extend the estimations of Rose (2000), also remark that "geography matters" and relate costs and benefits of a monetary union to neighborhood and the country size.
Non similar income levels (coefficient of variation of GDP$_{PPP}$s per capita) are often quoted as indicating similar preferences and therefore a high correlation of demand shocks (ρ_{demand}):

$$\rho_{demand} = \underset{(6.45)}{0.1} - \underset{(-5.16)}{0.25} \text{ coefficient of variation of GDP}_{per\ capita}.$$

Comparing the GDP$_{PPP}$s per capita in Table 2.7, p. 33 we can see that none of the CEE countries is in the range of the EU countries. Only the Greek part of Cyprus has a higher per capita income than the poorest EU countries Greece and Portugal. The top CEE countries are Slovenia and the Czech Republic.
A point emphasized by Bayoumi and Eichengreen (1994) is that in addition to the correlation of shocks their size and adjustment length should be taken into account. The similarity (coefficient of variation) of shock sizes and the adjustment speed reduces the costs of unification for a given correlation.
The identification procedure restricts the variance of the estimated shocks to unity. Like in Bayoumi and Eichengreen (1994) the shock size is measured by the associated impulse response functions. The size of supply shocks is measured by long-run effects, the size of demand shocks by the first-month impact on output and prices.
Table 2.10, p. 37 reveals high supply shocks for the CEE countries. Among them Hungary, Slovenia and Cyprus experienced the smallest shocks. Greece, Ireland and Luxembourg are the countries with the highest shocks among the EU. On average the EU11+4 group was subject to rather small shocks. In the lower part coefficients of variation of average national shock size were calculated for the EU and CEE groups, as well as for the groups identified in the shock correlation matrices. Unification costs (low coefficient of variation) are highest for the CEE+8 and less for EU+4, CEE5 and EU11.
The CEE countries experienced large demand shocks, among them Poland, Slovenia, and the Czech Republic had the smallest. The EU core group had the smallest shocks, besides the Netherlands, which has a size similar to the top CEE countries.

Both sizes of shocks give the group order EU11, EU+4, CEE5, and CEE+8. This result may be largely determined by the ongoing transition process.
In a theoretical paper on optimum currency areas Ricci (1997) derives the result that openness may have ambiguous effects on the costs and benefits of a currency union. He argues that openness may raise the costs since open economies are exposed to larger shocks. Figure 2.2, p. 36 shows a plot of the demand and supply shock size versus openness. There is no obvious relation between the shock size and openness. Adding GDP (in PPP) per capita as an explanatory variable gives the following regressions (without Cyprus):

$$\begin{aligned} \text{Supply shock size} &= \underset{(8.34)}{0.049} + \underset{(1.55)}{0.014}\,\text{Openness} - \underset{(6.87)}{0.0023}\,\text{GDP}_{PPP}\text{p.c.} \quad R^2 = 0.68 \\ \text{Demand shock size} &= \underset{(5.69)}{0.081} + \underset{(0.72)}{0.016}\,\text{Openness} - \underset{(5.49)}{0.0044}\,\text{GDP}_{PPP}\text{p.c.} \quad R^2 = 0.60 \end{aligned}$$

The shock size decreases significantly with the increasing income and increases insignificantly with openness.
The speed of adjustment to shocks in Table 2.12, p. 39 is measured by the response after two months as a share of the long-run effect. It is interesting to see that the CEE+8 countries adjust very fast. The speed of adjustment might be higher in transition countries because they are engaged in a structural change anyway. This is in line with the Olson hypothesis that long political stability (high development) causes sclerosis (low adjustment speed).
The similarity of the adjustment speed which favors unification is highest in EU11 and lower in EU+4, CEE5 and CEE+8.
In a structural setting the shock size and adjustment speed cost factors can be balanced against the shock correlation by relating them to a welfare measure.

2.2.2.2.1 Different productivity growth rates induce potential costs to a monetary union. This can be illustrated in a simple relation between wage, price and productivity growth $(\hat{w}, \hat{p}, \hat{q})$

$$\hat{p}_i = \hat{w}_i - \hat{q}_i \text{ wage equation country } i \tag{2.8}$$

$$\hat{p}_j = \hat{w}_j - \hat{q}_j \text{ wage equation country } j \tag{2.9}$$

$$\hat{e} = \hat{p}_i - \hat{p}_j \text{ equilibrium in goods market.} \tag{2.10}$$

Equation (2.10) is an equilibrium condition formulated as the purchasing power parity. $\hat{e}$ is a devaluation rate of country i's currency against currency j. It specifies that if country i experiences a higher inflation rate, its currency has to devalue in order to maintain price competitiveness. In a monetary union the nominal exchange rate cannot change ($\hat{e} = 0$) and therefore inflation rates have to be equal. A higher inflation rate in country i erodes its price competitiveness and growth. This gives the condition for a monetary union, where none of the members lose or win price competitiveness: $\hat{p}_i = \hat{p}_j$. Combining this condition with (2.8) and (2.9) gives: $\hat{w}_i - \hat{q}_i = \hat{w}_j - \hat{q}_j$ or $\hat{w}_i - \hat{w}_j = \hat{q}_i - \hat{q}_j$. If country i

has a lower productivity growth than country j, it's nominal wage growth must be lower in order to maintain price competitiveness. Therefore the effects of a monetary union for each country depend on the reaction of wage negotiations on shocks. This underpins the importance of the next paragraph.

2.2.2.2.2 Different labor market institutions. The effects of a supply shock might be influenced by the degree of wage bargaining centralization. The ILO mentions in their World Employment Report 1996-97 that "there is a clear trend towards greater independence of enterprizes and more individualized labor relationships." Since in a monetary union the nominal exchange rate is not available for adjusting relative wages, countries with sectoral or national wage bargaining structure might have greater difficulties to balance the missing nominal exchange rate by wage flexibility. Bargaining at the company level may be more flexible to cope with shocks.
In Table 2.14, p. 40 a clear tendency is that the former socialist countries except Slovenia, Ireland and the UK, and partly Bulgaria and Finland, do wage bargaining at the company level indicating that they may exhibit more flexible labor markets.

2.2.2.3 Economic policy shocks

The need for real exchange rate adjustment might also be triggered by economic policy shocks. It is clear that an ongoing transformation of an economy is a severe economic policy shock and that the transformations induce the EU countries to adjust, too. For example, countries with labor intensive production structures in the EU might need real exchange rate adjustment relative to capital intensive producing EU countries, if the eastern European countries produce and export relatively more labor intensive goods.

2.2.2.3.1 Transformation process. The development of the transformation process of the CEE countries is a clue to the expectations about future economic policy shocks and therefore the need for real exchange rate adjustment.
The development of the transformation processes is summarized in Table 2.15, p. 41 by some important aspects of the economies in 1998: *political stability* concerns the actual political situation, *currency* values stability and convertibility, *privatization* ranks the privatization success, *legal system* rates the development in the areas of economy, taxes, and administration and considers implementation capabilities, *banking sector* classifies the development of the private banking sector, the central bank and the possibilities for foreign banks to enter the markets, *infrastructure* appraises the existence of national and international transportation possibilities and *economic development* is an overall judgement.

The indicators show Poland, Hungary, the Czech Republic and Estonia as the top developed and Bulgaria and Romania as rather badly developed countries. Thus future shocks in the bottom (top) countries are expected to be high (low).

2.2.2.3.2 Deficits imply real appreciation, particularly if they are financed from abroad and spent in the domestic markets. Table 2.16, p. 42 shows the fiscal performance.
The best performance among eastern countries can be attested to Estonia, the Czech Republic, Slovenia and Poland. Hungary has a rather poor fiscal performance and distorts the group order in the lower part of the table. Excluding Hungary gives a -0.5% deficit for CEE5 which then is by far the top group.

2.2.2.3.3 Seignorage can be viewed as an inflation tax levied by raising the monetary base. Entering the EMU endangers the usage of this revenue, since profits of the ECB are shared among the member countries and a country specific seignorage is out of reach. However, the loss of seignorage determination can be viewed as a cost factor if one accepts money as an object within an optimal taxation context. One can argue that money is an input good and therefore should not be taxed. On the other hand, if no interest is paid on money then already price stability taxes money holdings.
For transition countries it is maybe more important to note that the inflation tax can be viewed as taxation of a good that shows low price elasticity due to a high degree of the black-market which cannot be taxed otherwise. In the CEE countries these activities account for up to a third of GDP.[11] According to an unpublished report of the European Commission this share amounts up to 15% in the EU.[12] After losing the inflation tax a shock by the need for higher tax rates can be expected.
Table 2.17, p. 43 reports seignorage income as a share of GDP. The group order shows that the EU countries rely less on the monetary monopoly revenues than the CEE countries.

2.2.3 Alternative adjustment instruments

2.2.3.1 Wage elasticity of labor supply

One alternative instrument for nominal exchange rate changes is a high real wage elasticity of labor supply. This can be best illustrated by the extreme case of an infinite elasticity and a horizontal supply function. In such a case there will be no real exchange rate changes since they would imply real wage changes which in turn would be eliminated in case of an infinite elasticity. Therefore nominal exchange rate changes are not necessary. The wage bargaining structure can

[11] FAZ GmbH Informationsdienste (1998)
[12] Eco (1998)

be seen as an adjustment instrument, too. An international wage elasticity of labor supply is the measure to calculate. Without appropriate data a proxy has to be constructed: Variation in the international unemployment rates is an index for mobility by the assumption that unemployment initiates migration until unemployment rates are equalized. High variance shows low mobility. In Table 2.18, p. 44 the international variation is calculated as the ratio between the highest and lowest national unemployment rates and its coefficient of variation. Both measures show lower mobility in the EU than in the USA, where 51 state unemployment rates were used. The US unemployment variation is smaller for all group combinations.
It can be expected that international mobility is mainly determined by barriers like administrative and linguistic relocation costs. Table 2.19, p. 44 shows that the national variation is much lower than international.
Mobility is likely to vary with different incomes and therefore may not cover the whole range of wages as a substitute for nominal exchange rate changes.
Summarizing the findings, labor mobility cannot be expected to substitute nominal exchange rate changes either in the EU or in the CEE countries, especially as nominal exchange rate changes adjust for international imbalances and international mobility seems to be underdeveloped in Europe.

2.2.3.2 Capital mobility

The substitution of nominal exchange rate changes by capital mobility has to be investigated for supply and demand shocks separately.
For instance, since the EU produces rather capital intensive goods compared to the CEE countries, the CEE demand for these goods induces an expansionary demand shock to the capital intensive producing countries, like Germany. The primary effect is a real appreciation of the Euro and a subsequent capital import which implies further appreciation. The interest payments to foreigners dampen this appreciation.
In the case of an expansionary supply shock, for instance through rising productivity, the primary effect is a real depreciation, which is dampened by capital imports due to higher profits.
We have a reinforcing effect of capital mobility in case of a demand shock but a compensating effect in case of a supply shock. The suited measure is a shock elasticity of capital flows.
A possible proxy are factor payments and receipts entering the national accounts. This includes employees' and entrepreneurship income as well as asset income. These figures show international migration as well as capital movements. Table 2.20, p. 45 shows not only factor receipts as a share of GDP, which reflect the natives' mobility but also the factor payments which point towards favorable conditions for factor receipts.

In interpreting the figures as proxies for movements we see that the CEE countries fall behind compared to the EU countries, except for Malta, Hungary, and Estonia.

2.2.3.3 Fiscal transfers

Another possible substitute for nominal exchange rate changes are fiscal transfers. The instrument's effectiveness as a built-in-stabilizer has also to be distinguished in case of a demand and supply shock.
An expansive demand shock is followed by a real appreciation. It also reduces transfers to the country Tr in X-Im+Tr$\equiv \Delta$NFA implying a real depreciation and therefore a compensating effect. In case of a positive supply shock a real depreciation sets in. Due to a growing prosperity, transfers imply a further depreciation and therefore reinforce the primary effect.
However, the possibilities of this alternative instrument are not large. One reason is that the EU budget is relatively small. Its size is around 2% of all government disbursements. The other reason is that the EU's transfers are not constructed flexibly enough to cope with shocks.[13] In relation to the CEE countries that there will not be massive transfers in case of shocks since in the accession talks the old members have limited transfers to the CEE countries.
As a result it can be stated that fiscal transfers cannot be expected to be an effective alternative instrument in the future.

2.2.4 Criticisms of cost criteria

There are two ways of criticizing these possible objections to monetary union: first, the macroeconomic costs of a monetary union are negligible since nominal exchange rate changes are not necessary. Because of unimportant differences between the countries (section 2.2.4.1) or because of the ineffectiveness of the nominal exchange rate instrument (section 2.2.4.2) and second, the benefits of a monetary union are stressed (section 2.3).

2.2.4.1 Reduction of asymmetric shocks

The historical record of shocks is probably a bad predictor for future shocks. Even more since the shock correlation itself may be influenced by the establishment of a monetary union. If a common currency increases trade among union members their structures converge and shocks become more and more symmetric. Therefore, a historical shock correlation criterion overestimates the costs. The Krugman-Eichengreen hypothesis formulates the opposite effect, where each country is expected to specialize according to its comparative advantages.[14] Em-

[13] Heinemann (1995)
[14] Krugman (1993), Eichengreen (1992)

pirical findings strongly oppose the Krugman-Eichengreen effect but still pointing towards the endogeneity of the optimal currency area criteria.[15]
Becoming a member of a monetary union raises the suitability for a monetary union. Frankel and Rose (1998) test this hypothesis by explaining the business cycle correlation by trade intensity (without identifying supply and demand shocks): "the more one country trades with others, the more highly correlated will be their business cycles".
Another endogeneity test is to explain the trade specialization by trade intensity. For this purpose bilateral trade flows are used. They are extracted from Eurostat's COMEXT database for the EU11+4 countries and 100 commodity categories in 1995. Import and export specializations are explained by trade intensity in two separate regressions. Trade intensities between countries j and i are defined as export (import) from country j to country i as a share of the sum of all EU exports (imports) of countries j and i.
The trade intensity may simultaneously be determined by trade specialization. For instance, if Germany needs broadly diversified commodities from France, it is likely that this import flow is high, compared to other German imports. Therefore trade intensity is instrumentized by the gravity model variables distance between main industrial regions, a dummy for a common border and a dummy for a common language. The first stage estimation yields highly significant coefficients with R^2s around 0.7. The simple OLS and two stage instrument (IV) estimations are as follows (t-statistics in parenthesis):

$$\begin{aligned}
\text{OLS: Import specialization} &= \underset{(85.77)}{0.70} - \underset{(-7.00)}{1.34}\ \text{Import Intensity} \\
\text{OLS: Export specialization} &= \underset{(81.79)}{0.70} - \underset{(-7.40)}{1.53}\ \text{Export Intensity} \\
\text{IV: Import specialization} &= \underset{(78.19)}{0.71} - \underset{(-7.95)}{1.86}\ \text{Import Intensity} \\
\text{IV: Export specialization} &= \underset{(73.85)}{0.72} - \underset{(-7.96)}{2.05}\ \text{Export Intensity.}
\end{aligned}$$

Trade integration decreases trade specialization (increases trade diversification) highly significantly. A trade intensifying common currency as in Frankel and Rose (1999) increases trade diversification and thus a monetary union reduces the need for exchange rate changes in case of shocks. Therefore, as in the paper of Frankel and Rose (1998), the optimum currency area criterion is endogenous.

2.2.4.2 Ineffectiveness of the nominal exchange rate instrument

This criticism questions the possibility of a durable real depreciation through nominal exchange rates and therefore the abandonment of nominal exchange rate changes induces no costs.

[15] Frankel and Rose (1998)

In the case of a negative demand shock the equilibrating adequate reaction is a real depreciation in order to shift the demand curve to the right. The criticism argues: a depreciation raises import prices of input goods and shifts the supply curve up, moreover a depreciation reduces the real wages. Whenever wage bargaining reacts by higher nominal wages, the supply curve shifts up again. In the end both effects on the supply curve compensate the intended quantity effect of the depreciation. The result is merely a price effect. Summarizing the argument: in case of a wage-price spiral the intended effect of a depreciation on price competitiveness is undone by rising prices.
The counter argument is that the wage-price spiral is not set in place immediately. Therefore the question reduces to: how long does it take before the effect of a depreciation on price competitiveness is undone? One can argue that a depreciation is a rather low-cost instrument compared to the adjustment of prices and wages. This is even more true in the case of rigid labor markets.
A historical example is the German unification, where an expansionary fiscal policy shock created inflation pressure, which was met by a more restrictive monetary policy which in turn spilled over by the ERM to other EU countries. A better idea would have been a DM appreciation, which would have decreased import prices and dampened the inflationary pressure. Rising interest rates, which contributed to the recession, would not have been necessary.

2.2.4.3 Devaluation and credibility of anti-inflationary policy

This argument claims that flexible nominal exchange rates involve more costs than benefits. It runs analogous to the Barro-Gordon model and states that an exchange rate instrument discretionarily used by politicians leads to a prisoner's dilemma with a Pareto inefficient Nash-equilibrium. Therefore an anti-inflationary policy and a simultaneously discretionary exchange rate instrument are not credible. The implication is that a monetary union implies an import of credibility and reduces the costs of keeping the inflation low. Countries with high inflation (Table 2.21, p. 46) profit from a monetary union.
A premise for this argument is of course that the ECB has a higher credibility. A less complicated possibility for a credible monetary policy is an independent central bank rather than a monetary union.

2.3 Benefits of a monetary union

Costs of a monetary union arise mainly at the macro level whereas benefits arise rather at the micro level.
The most apparent benefits are superfluous exchange transaction costs. In addition to these direct benefits, an indirect benefit is the reduced scope for price

discrimination and thus stronger competition. Further benefits are reduced uncertainties about exchange rates and resulting allocation improvements.
All these benefits refer to foreign trade and capital movements which can be approximated by trade integration as measured in Table 2.22, p. 46 by the share of exports and imports in GDP. Small European countries would profit most from a monetary union since they are most open.

2.4 Results

Table 2.23, p. 47 summarizes the results with pluses indicating suitability of the EMU according to some criteria. The row "Weights real exchange rate changes" weights the pluses of the criteria except for the real exchange rate criterion. "Weights" shows the overall weights. These weights are of course subject to criticism as any normative judgement. They cannot be eliminated as long as the OCA theory is a collection of criteria approximating the social welfare function which is not explicitly formulated within a model of the involved variables.
For the shock correlation results column "CC or RE or LE" is added, disqualifying countries in the peripheral group in the demand shock correlation, since the core group is correlated with the R,L and E but not the P group.
The column "Suitability index" calculates the result. Qualitatively similar orders were calculated for real exchange rate criterion weights between 0 and 1.
The weights 0.05, 0.1 and 0.2 capture the main characteristics: Denmark, the Netherlands, Austria, Belgium, Germany, Luxembourg and France are always at the top of suitability. Then all other EU11+4 countries except for Italy follow: .., France, Finland, Spain, Sweden, Ireland, the United Kingdom, Portugal, Greece. The least suitable CEE country order is unchanged for all three weights as in Table 2.23: Hungary, Turkey, the Slovak Republic, Bulgaria, Latvia, Romania, Lithuania.
The top CEE country order Cyprus, Malta, Slovenia, Estonia, the Czech Republic, and Poland is also unaffected by the weight, except for Malta which profits from a higher weight. With a lower weight the top CEE countries Cyprus and Slovenia rank before Portugal and the Czech Republic, Poland, and Estonia before Greece.
The real exchange rate criterion takes into account events from as early as the beginning of the nineties, as opposed to the other criteria which rather based on figures from the end of the decade. Therefore its weight can be interpreted as a backward looking parameter. If one looks less at the past, lowering the weight, Cyprus, Slovenia, the Czech Republic and Poland emerge as the most suitable candidates for the EMU.
Two additional points regarding the Greek EMU-Enlargement can be noticed: the real exchange rate changes of Greece in the years 1996-1998 do not raise the EU11 standard deviation $\sigma_{R,\,EU11} \geq \sigma_{R,\,EU11+Greece}$, Greece's supply shocks

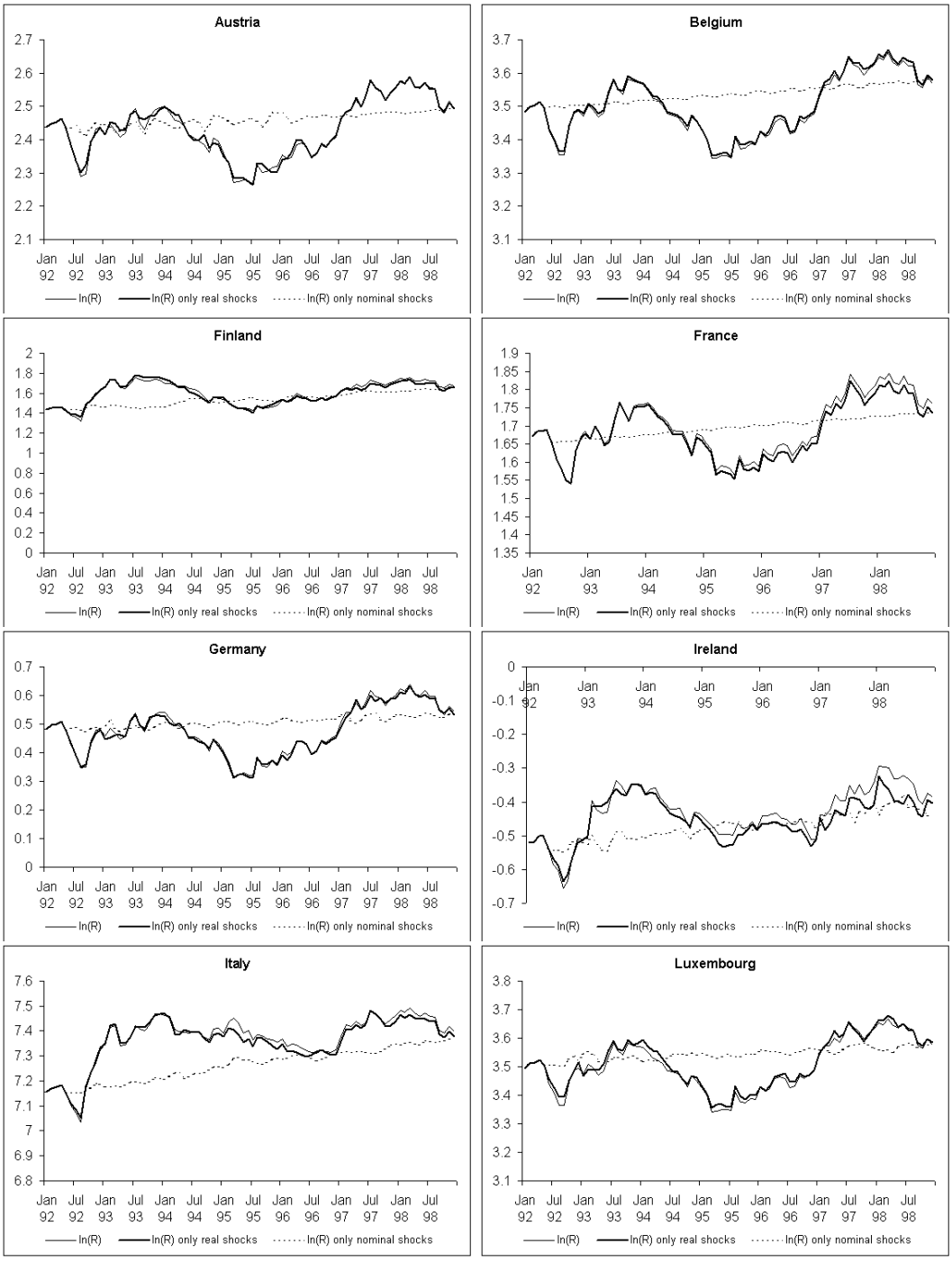
are positively correlated with those of France, Belgium, Ireland and Portugal but negatively with those of all other EU11 countries. Its demand shocks are negatively correlated only with those of Belgium.

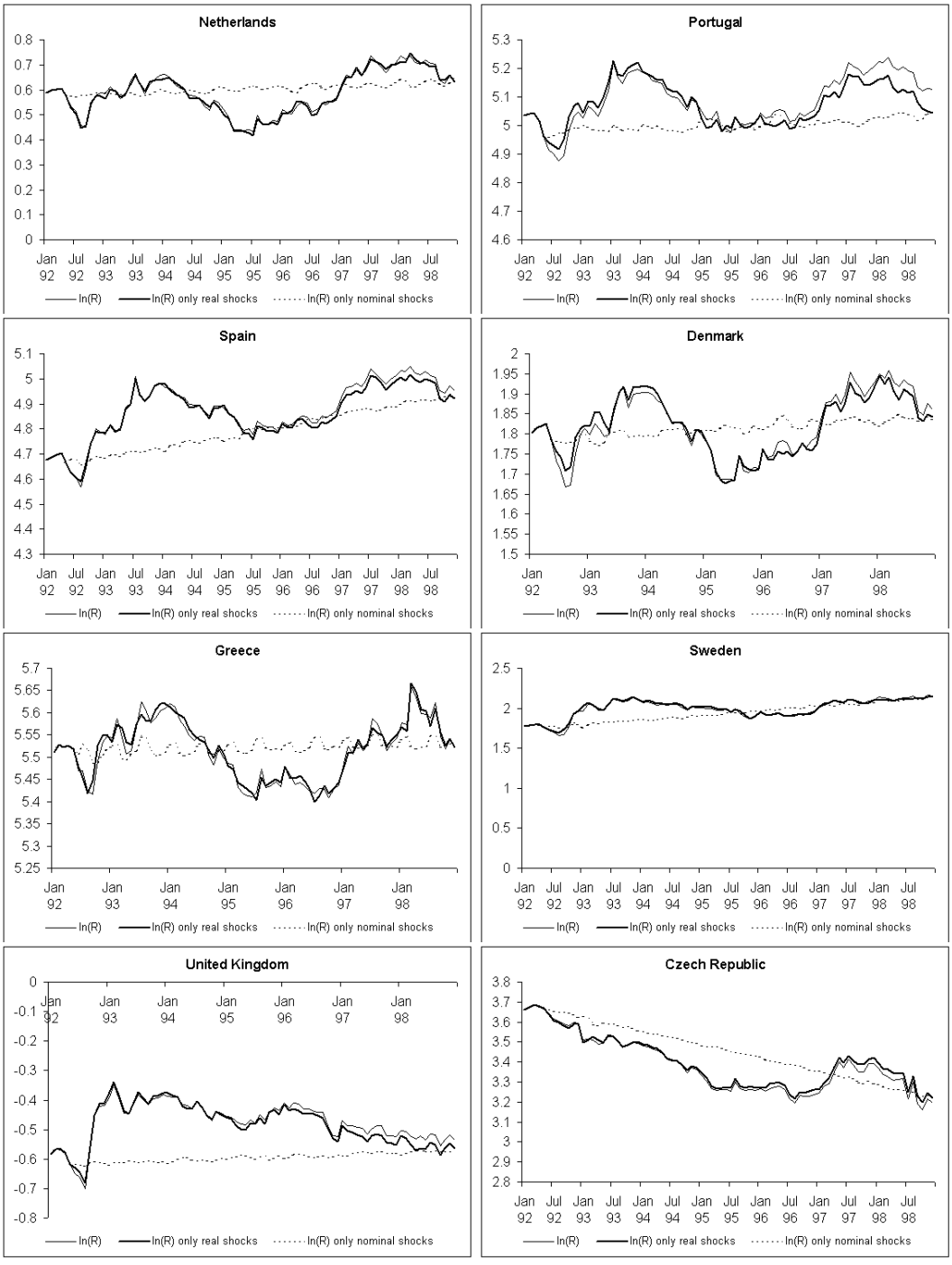
2.5 Conclusion

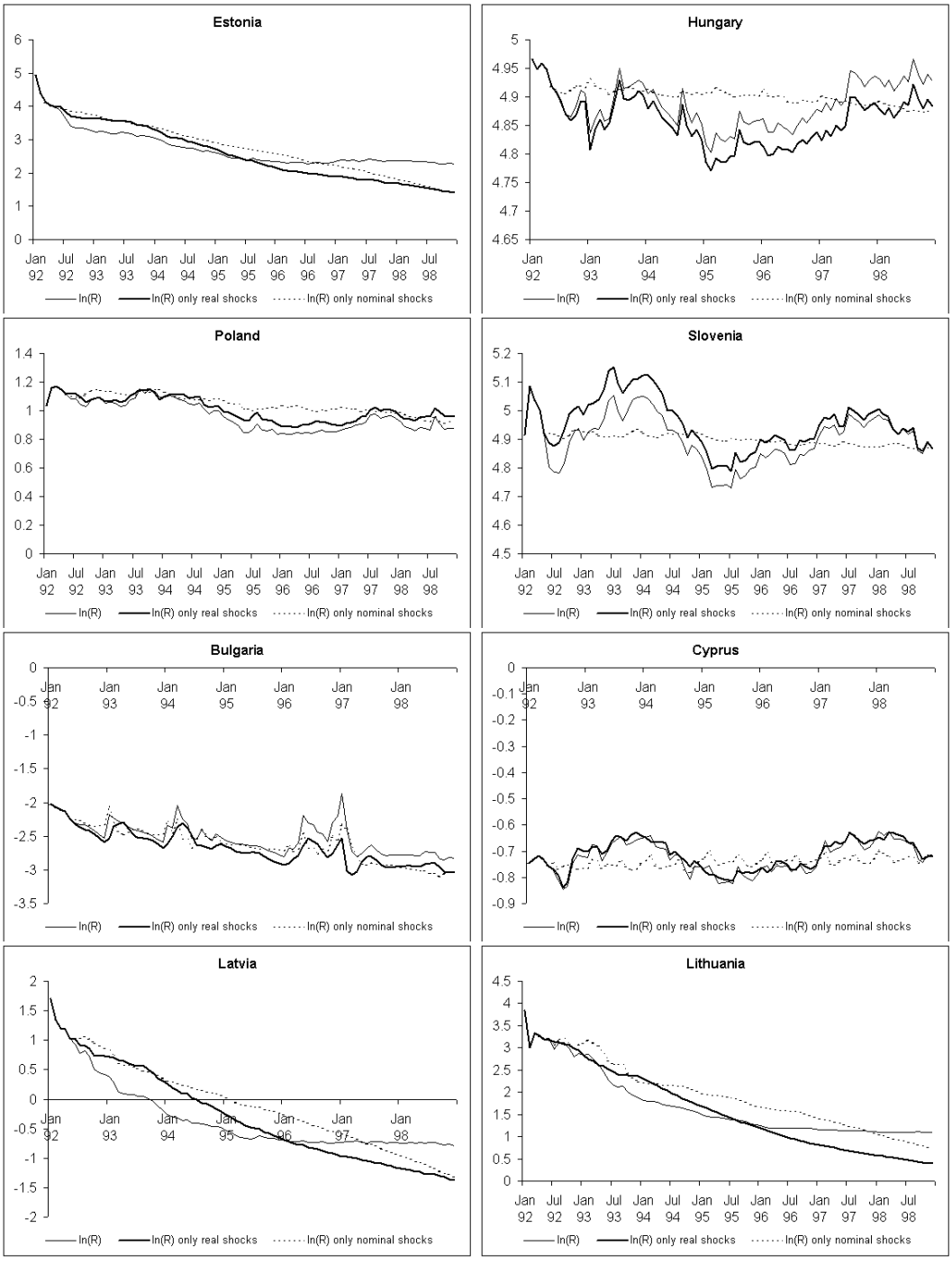
This traditional empirical approach to the objective of the optimum currency area by evaluating some criteria is seriously incomplete.
The first is the endogeneity problem and comes as a DM zone for the core European countries, where the causal direction between optimum currency area and de facto currency area is unclear. Trade effects of a currency union and nominal shocks on the real exchange rate are not accounted for in the Vaubel criterion, too, as he indicates himself. All these problems are due to the neglected Lucas critique and the absence of a structural model.
The second problem is the disconnection of the OCA criteria from a social welfare function. A quantification of the welfare costs and benefits of a currency union in absolute terms in contrast to the mere suitability ordering is impossible. This is related to the absence of an agreed weighting of the criteria. Again the only way to obtain an (implicit) weighting would be to formulate and estimate explicit assumptions on the relationships between the underlying variables and welfare.
The OCA objective consists of a monetary policy rule evaluation where policy rules represent different exchange rate systems. In Taylor (1993) or more up to date in Rotemberg and Woodford (1998) an empirical framework is presented based on rational expectations for the evaluation of different monetary rules in a multi country setting. Taylor used such an approach to evaluate a social welfare function (objective) over different exchange rate regimes. The same approach is needed for a welfare evaluation over different groups of countries. In the light of such an approach the traditional OCA criteria can be viewed as proxies for an implicit social welfare function.

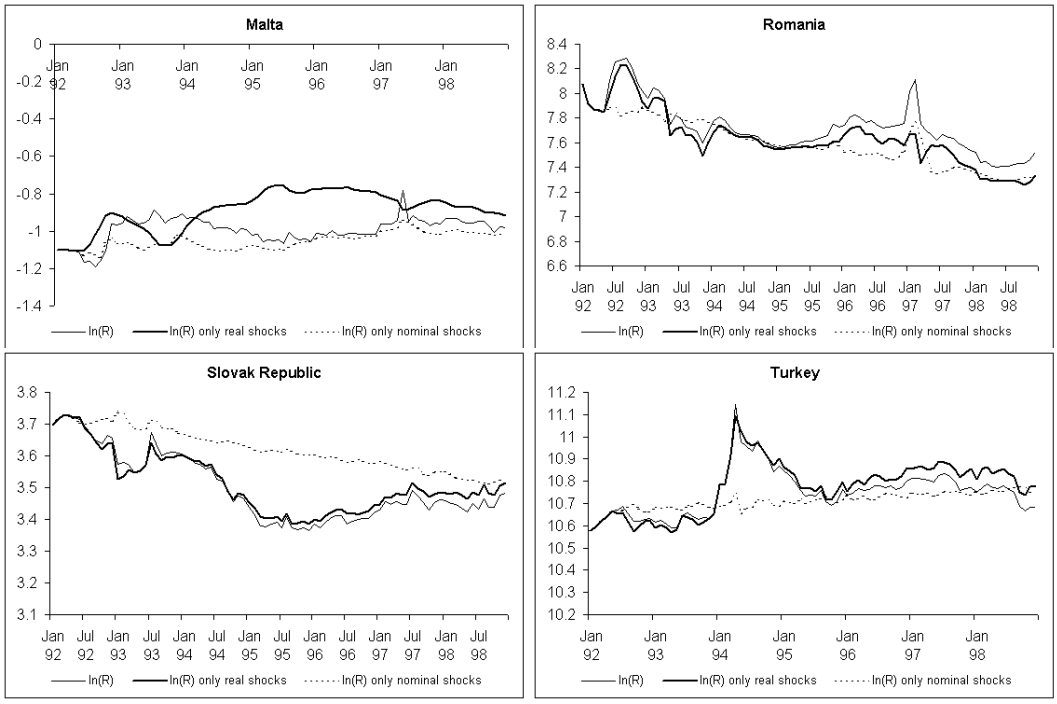
2.6 Appendix

2.6.1 Decomposed real exchange rates




Netherlands
Portugal
Spain
Denmark
Greece
Sweden
United Kingdom
Czech Republic
ln(R)
ln(R) only real shocks
ln(R) only nominal shocks


Estonia
Hungary
Poland
Slovenia
Bulgaria
Cyprus
Latvia
Lithuania
ln(R)
ln(R) only real shocks
ln(R) only nominal shocks


Malta
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
Jan 92
Jan 93
Jan 94
Jan 95
Jan 96
Jan 97
Jan 98
ln(R)
ln(R) only real shocks
ln(R) only nominal shocks
Romania
8.4
8.2
8
7.8
7.6
7.4
7.2
7
6.8
6.6
Jan 92
Jul 92
Jan 93
Jul 93
Jan 94
Jul 94
Jan 95
Jul 95
Jan 96
Jul 96
Jan 97
Jul 97
Jan 98
Jul 98
ln(R)
ln(R) only real shocks
ln(R) only nominal shocks
Slovak Republic
3.8
3.7
3.6
3.5
3.4
3.3
3.2
3.1
Jan 92
Jul 92
Jan 93
Jul 93
Jan 94
Jul 94
Jan 95
Jul 95
Jan 96
Jul 96
Jan 97
Jul 97
Jan 98
Jul 98
ln(R)
ln(R) only real shocks
ln(R) only nominal shocks
Turkey
11.2
11.1
11
10.9
10.8
10.7
10.6
10.5
10.4
10.3
10.2
Jan 92
Jan 93
Jan 94
Jan 95
Jan 96
Jan 97
Jan 98
ln(R)
ln(R) only real shocks
ln(R) only nominal shocks

2.6.2 Tables and Figures

Table 2.1: Real Exchange Rate Changes 1992-1998

7 Years, 1992-98					
EU11	Spain	-1.21	CEE+8	Malta	1.25
EU+4	Sweden	-0.99	CEE+8	Cyprus	1.88
EU11	Italy	-0.30	CEE5	Hungary	2.35
EU11	Finland	0.39	EU+4	Greece	2.40
EU11	Portugal	0.40	EU+4	United Kgd.	2.92
EU11	France	0.52	CEE5	Slovenia	3.13
EU11	Ireland	0.53	CEE5	Poland	6.70
EU11	Luxembourg	0.54	CEE+8	Slovak Rep.	6.71
EU11	Belgium	0.58	CEE5	Czech Rep.	8.39
EU11	Germany	0.72	CEE+8	Romania	10.16
EU11	Austria	0.82	CEE+8	Bulgaria	10.75
CEE+8	Turkey	0.83	CEE5	Estonia	29.45
EU11	Netherlands	0.85	CEE+8	Latvia	31.28
EU+4	Denmark	1.03	CEE+8	Lithuania	41.14
Standard deviations					
EU11		0.57			
EU+4		1.51			
CEE5		9.97			
CEE+8		14.09			

Source: IMF

Table 2.2: Multiples of the Standard Deviation of Real Exchange Rate Changes

Years	Start	End	EU11	EU+4	EU11+4	CEE5	CEE+8
7	1992	1998	$0.9_{0.6}$	$2.5_{1.5}$	$1.7_{1.0}$	$16.6_{10.0}$	$23.5_{14.1}$
5	1994	1998	$1.8_{1.1}$	$3.8_{2.3}$	$2.10_{1.8}$	$7.3_{4.4}$	$9.0_{5.4}$
3	1996	1998	$1.6_{0.9}$	$8.3_{5.0}$	$4.8_{2.9}$	$2.3_{1.4}$	$12.8_{7.7}$
3	1993	1995	$4.7_{2.8}$	$2.5_{1.5}$	$4.3_{2.6}$	$23.5_{14.1}$	$34.2_{20.5}$
1	1998	1998	$1.7_{1.0}$	$9.1_{5.5}$	$6.3_{3.8}$	$13.8_{8.3}$	$8.10_{5.4}$
1	1997	1997	$1.6_{1.0}$	$6.6_{4.0}$	$5.2_{3.1}$	$7.6_{4.6}$	$44.7_{26.8}$
1	1996	1996	$6.7_{4.0}$	$9.5_{5.7}$	$8.4_{5.0}$	$5.1_{3.1}$	$20.6_{12.4}$
1	1995	1995	$4.5_{2.7}$	$9.10_{6.0}$	$6.5_{3.9}$	$16.4_{9.8}$	$19.3_{11.6}$
1	1994	1994	$5.8_{3.5}$	$4.9_{2.9}$	$5.7_{3.4}$	$23.4_{14.0}$	$19.5_{11.7}$
1	1993	1993	$7.3_{4.4}$	$8.8_{5.3}$	$7.9_{4.7}$	$18.7_{11.2}$	$105.2_{63.1}$
1	1992	1992	$13.3_{8.0}$	$13.4_{8.1}$	$13.9_{8.3}$	$226.1_{135.7}$	$128.2_{76.9}$

Years	Start	End	EU11 CEE5	EU11 CEE+8	EU11+4 CEE5	EU11+4 CEE+8	USA11
7	1992	1998	$11.9_{7.2}$	$18.5_{11.1}$	$10.9_{6.5}$	$17.0_{10.2}$	$0.7_{0.4}$
5	1994	1998	$6.4_{3.8}$	$9.0_{5.4}$	$5.9_{3.5}$	$8.4_{5.0}$	$0.7_{0.4}$
3	1996	1998	$4.0_{2.4}$	$12.1_{7.3}$	$5.1_{3.0}$	$11.6_{7.0}$	$0.8_{0.5}$
3	1993	1995	$15.5_{9.3}$	$24.6_{14.8}$	$14.5_{8.7}$	$22.7_{13.6}$	$1.0_{0.6}$
1	1998	1998	$7.9_{4.8}$	$7.8_{4.7}$	$7.8_{4.7}$	$8.2_{4.9}$	$1.3_{0.8}$
1	1997	1997	$4.6_{2.7}$	$35.1_{21.1}$	$5.7_{3.4}$	$32.2_{19.3}$	$1.5_{0.9}$
1	1996	1996	$6.6_{3.9}$	$14.3_{8.6}$	$7.6_{4.6}$	$14.0_{8.4}$	$1.0_{0.6}$
1	1995	1995	$10.3_{6.2}$	$13.7_{8.2}$	$10.5_{6.3}$	$13.1_{7.9}$	$1.0_{0.6}$
1	1994	1994	$15.3_{9.2}$	$14.7_{8.8}$	$14.5_{8.7}$	$13.9_{8.3}$	$1.5_{0.9}$
1	1993	1993	$13.6_{8.2}$	$80.7_{48.4}$	$12.6_{7.5}$	$74.5_{44.7}$	$2.0_{1.2}$
1	1992	1992	$139.9_{83.9}$	$90.7_{54.4}$	$130.0_{78.0}$	$83.9_{50.3}$	$1.0_{0.6}$

Source: IMF; Nummeraire is USA11 1992. Subindexes are the standard deviations.

Table 2.3: Simulated Country Inflation Rates for a Zero Inflation Monetary Union 1992-1998

	EU11	EU+4	EU11+4	CEE+8	CEE5	EU11+CEE5	EU11+CEE+8	EU11+4+CEE5	EU11+4+CEE+8
Netherlands	0.4		0.1			-0.6	-0.7	-0.5	-0.7
Austria	0.3		0.1			-0.6	-0.7	-0.6	-0.7
Germany	0.3		0.1			-0.6	-0.7	-0.6	-0.7
Belgium	0.2		0.0			-0.6	-0.8	-0.6	-0.7
Luxembourg	0.1		0.0			-0.6	-0.8	-0.6	-0.7
Ireland	0.1		0.0			-0.6	-0.8	-0.6	-0.7
France	0.1		-0.1			-0.7	-0.8	-0.6	-0.7
Portugal	0.0		-0.1			-0.7	-0.8	-0.6	-0.8
Finland	0.0		-0.1			-0.7	-0.8	-0.6	-0.8
Italy	-0.5		-0.6			-0.8	-0.9	-0.8	-0.9
Spain	-1.2		-1.1			-1.0	-1.0	-1.1	-1.0
United Kingdom		0.7	1.4					0.0	-0.3
Greece		0.5	1.1					-0.1	-0.4
Denmark		-0.1	0.3					-0.5	-0.7
Sweden		-1.0	-1.0					-1.0	-1.0
Estonia				1.8		6.0		6.7	
Czech Republic				-0.1		1.2		1.4	
Poland				-0.3		0.8		0.9	
Slovenia				-0.6		-0.1		0.0	
Hungary				-0.7		-0.2		-0.2	
Turkey					-0.9		-0.7		-0.7
Slovak Republic					-0.4		0.2		0.3
Romania					-0.2		0.7		0.9
Malta					-0.8		-0.7		-0.6
Lithuania					2.0		5.3		6.1
Latvia					1.3		3.8		4.5
Cyprus					-0.8		-0.6		-0.5
Bulgaria					-0.2		0.8		1.0

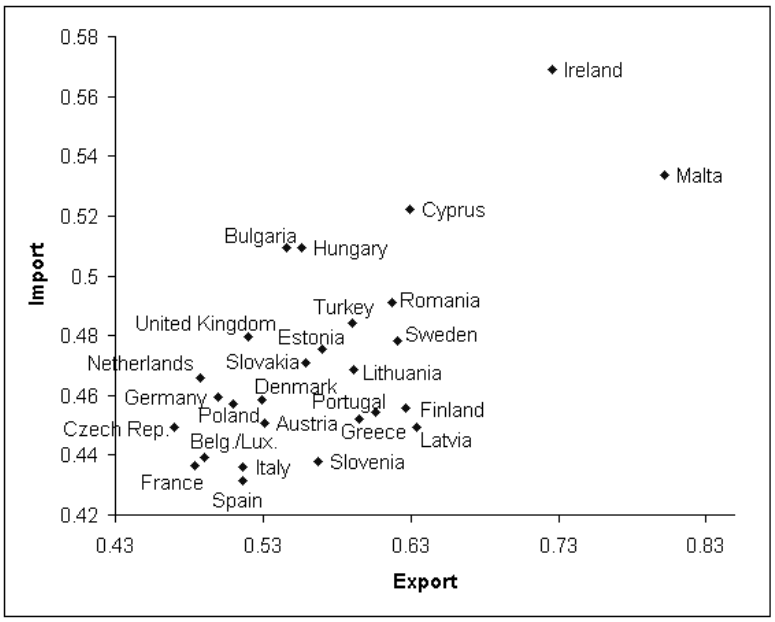
Source: IMF

Table 2.4: Minimizing the Standard Deviation of Real Exchange Rate Changes

7 Years, 1992-98				5 Years, 1994-98				3 Years, 1996-98			
All		EU11+		All		EU11+		All		EU11+	
10.56		10.56		5.13		5.13		6.47		6.47	
8.12	Lithuania	8.12	Lithuania	4.26	Lithuania	4.26	Lithuania	5.30	Romania	5.30	Romania
6.24	Latvia	6.24	Latvia	3.74	Estonia	3.74	Estonia	3.76	Bulgaria	3.76	Bulgaria
3.36	Estonia	3.36	Estonia	3.29	Latvia	3.29	Latvia	3.42	Lithuania	3.42	Lithuania
2.94	Bulgaria	2.94	Bulgaria	2.74	Bulgaria	2.74	Bulgaria	3.00	UK	3.00	UK
2.45	Romania	2.45	Romania	2.49	Romania	2.49	Romania	2.75	Turkey	2.75	Turkey
2.02	Czech Rep.	2.02	Czech Rep.	2.30	Turkey	2.30	Turkey	2.46	Latvia	2.46	Latvia
1.68	Slovak Rep.	1.68	Slovak Rep.	2.10	Poland	2.10	Poland	2.25	Estonia	2.25	Estonia
1.15	Poland	1.15	Poland	1.85	UK	1.85	UK	2.04	Czech Rep.	2.04	Czech Rep.
1.05	Slovenia	1.05	Slovenia	1.52	Czech Rep.	1.52	Czech Rep.	1.92	Sweden	1.92	Sweden
0.95	UK	0.95	UK	1.27	Slovak Rep.	1.27	Slovak Rep.	1.82	Malta	1.82	Malta
0.84	Spain	0.87	Greece	1.22	Slovenia	1.22	Slovenia	1.68	Poland	1.68	Poland
0.73	Sweden	0.77	Hungary	1.16	Malta	1.16	Malta	1.51	Slovak Rep.	1.51	Slovak Rep.
0.63	Greece	0.69	Sweden	1.07	Italy	1.10	Greece	1.31	Slovenia	1.31	Slovenia
0.49	Hungary	0.61	Cyprus	0.98	Greece	1.07	Cyprus	1.15	Cyprus	1.15	Cyprus
0.37	Cyprus	0.60	Malta	0.91	Cyprus	1.03	Hungary	0.93	Hungary	0.93	Hungary
0.26	Italy	0.59	Denmark	0.80	Hungary	1.03	**Denmark**	0.80	Ireland	0.93	**Denmark**
0.20	*Malta*	0.59	**Turkey**	0.68	Ireland	1.03	**Sweden**	0.57	Italy	0.93	**Greece**
0.17	**Denmark**		Austria	0.60	Portugal		Austria	0.52	Greece		Austria
0.15	**Netherlds.**		Belgium	0.55	Sweden		Belgium	0.44	Portugal		Belgium
0.13	**Turkey**		Finland	0.47	Denmark		Finland	0.26	Denmark		Finland
0.11	**Austria**		France	0.38	Spain		France	0.24	*Netherlands*		France
0.08	**Germany**		Germany	0.30	Finland		Germany	0.22	*Luxembourg*		Germany
0.07	**Finland**		Ireland	0.20	France		Ireland	0.19	*Austria*		Ireland
0.02	**Portugal**		Italy	0.11	*Netherlands*		Italy	0.09	**Belgium**		Italy
0.01	**Belgium**		Luxembrg.	0.04	**Luxembrg.**		Luxembourg	0.02	**Germany**		Luxembourg
0.01	**France**		Netherlands	0.03	**Belgium**		Netherlands	0.00	**Finland**		Netherlands
	Ireland		Portugal		**Austria**		Portugal		**France**		Portugal
	Luxembrg.		Spain		**Germany**		Spain		**Spain**		Spain
0.41	USA11			0.41	USA11			0.47	USA11		
2.46	F USA / Ita			4.07	F USA / Fra			4.60	F USA / Lux		
0.08	p-val. (df 10,11)			0.09	p-val. (df 10,4)			0.05	p-val. (df 10,5)		
4.07	F USA / Mal			13.43	F USA / Net			6.13	F USA / Aus		
0.02	p-val. (df 10,10)			0.03	p-val. (df 10,3)			0.05	p-val. (df 10,4)		

Source: IMF

Figure 2.1: Trade Specialization 1998



Source: COMTRADE (UN)

Table 2.5: Average of Export and Import Specialization 1998

Czech Rep.	0.459	Greece	0.524
France	0.460	Bulgaria	0.528
Belg./Lux.	0.465	Lithuania	0.530
Spain	0.474	Portugal	0.530
Italy	0.476	Hungary	0.533
Netherlands	0.477	Turkey	0.537
Germany	0.479	Finland	0.541
Poland	0.483	Latvia	0.542
Austria	0.491	Sweden	0.550
Denmark	0.494	Romania	0.554
United Kgd.	0.500	Cyprus	0.576
Slovenia	0.503	Ireland	0.647
Slovakia	0.515	Malta	0.668
Estonia	0.523		

Source: COMTRADE (UN)

Table 2.6: Foreign Direct Investment, Net Inflows (% of GDP)

	1992	1993	1994	1995	1996	1997	Average
Hungary	3.97	6.09	2.76	10.12	4.39	4.55	5.31
Malta	1.46	2.28	5.67	5.64	9.66	3.85	4.76
Latvia	0.46	0.85	3.93	2.87	7.44	9.43	4.16
Estonia	1.94	4.04	5.31	4.21	3.44	5.69	4.10
Czech Rep.	3.94	2.10	2.20	5.05	2.54	2.47	3.05
Sweden	0.00	1.99	3.16	6.46	2.18	4.33	3.02
Ireland	2.77	2.29	1.55	2.24	3.70	3.63	2.70
Poland	0.80	1.99	2.03	3.08	3.34	3.62	2.48
Netherlands	2.43	2.73	2.23	2.89	1.96	2.42	2.44
United Kgd.	1.54	1.65	0.90	1.84	2.24	2.96	1.86
Denmark	0.69	1.22	3.31	2.29	0.42	1.70	1.61
Lithuania	0	0.50	0.54	1.05	1.93	3.70	1.54
Spain	2.30	1.70	1.94	1.11	1.11	1.04	1.53
France	1.65	1.66	1.19	1.55	1.43	1.65	1.52
Bulgaria	0.40	0.51	1.08	0.69	1.11	4.94	1.46
Portugal	1.98	1.83	1.44	0.65	0.65	1.68	1.37
Cyprus	1.56	1.28	1.04	1.41	..	..	1.33
Slovak Rep.	0	1.66	1.47	1.05	1.50	0.85	1.31
Romania	0.31	0.36	1.22	1.28	0.84	3.49	1.25
Austria	0.77	0.62	1.08	0.82	1.96	1.20	1.07
Finland	0.37	1.02	1.53	0.83	0.89	1.78	1.07
Slovenia	0.89	0.89	0.89	0.94	0.98	1.76	1.06
Greece	1.16	1.06	0.99	0.92	0.86	..	1.00
Turkey	0.53	0.35	0.47	0.52	0.40	0.42	0.45
Italy	0.32	0.44	0.21	0.45	0.29	0.32	0.34
Germany	0.13	0.10	0.09	0.56	-0.11	-0.02	0.13
Belg./Lux.	..	..	..	..	..	..	

Source: World development indicators (World Bank)

Table 2.7: GDP Sector Shares and GDP 1997

	SI	Agriculture	Industry	Services		GDP_{PPP} per capita in ,000$	GDP_{PPP} in million $	Population
Bulgaria	0.25	21	29	50	Romania	3.9	87.4	22.4
Romania	0.27	23	51	26	Latvia	4.2	9.8	2.4
Turkey	0.30	18	29	53	Bulgaria	4.3	34.9	7.8
Lithuania	0.37	10	32	58	Lithuania	4.8	17.3	3.6
Poland	0.43	5	35	60	Cyprus turkish	5.0	0.8	0.8
Czech Rep.	0.43	5	42	53	Estonia	5.6	7.9	1.4
Ireland	0.43	5	39	56	Turkey	6.2	409.4	65.7
Slovak Rep.	0.43	5	33	62	Poland	7.2	276.5	38.6
Slovenia	0.44	4	35	61	Hungary	7.8	79.4	10.1
Portugal	0.44	4	36	60	Slovak Rep.	8.5	45.9	5.4
Latvia	0.45	8	29	63	Slovenia	10.9	21.4	1.9
Finland	0.45	5	32	63	Czech Rep.	11.7	120.8	10.3
Spain	0.45	3	34	63	Malta	13.8	5.3	0.4
Greece	0.47	8	27	64	Greece	13.9	149.2	10.6
Hungary	0.48	5	30	65	Portugal	15.3	151.4	10.0
Estonia	0.49	4	31	66	Cyprus Greek	15.4	9.0	0.8
Italy	0.49	3	32	66	Spain	17.3	677.5	40.0
Austria	0.49	1	32	66	Ireland	20.3	73.7	3.8
Sweden	0.51	2	31	67	Sweden	20.7	184.0	8.9
Cyprus Turkish	0.52	12	21	68	Finland	21.0	108.6	5.2
Germany	0.53	1	30	68	Italy	21.4	1212.0	57.6
Denmark	0.54	4	27	69	United Kgd.	21.8	1290.0	59.5
Netherlands	0.55	4	27	70	Germany	22.7	1864.0	82.8
France	0.56	3	26	71	Netherlands	23.1	365.1	15.9
Malta	0.57	3	26	71	France	23.3	1373.0	59.3
Cyprus Greek	0.57	6	22	71	Austria	23.4	190.6	8.1
Belgium	0.57	1	27	72	Denmark	23.8	127.7	5.3
United Kgd.	0.60	2	25	73	Belgium	23.9	243.4	10.2
Luxembourg	0.64	1	23	76	Luxembourg	34.2	14.7	0.4

Source: World Fact Book (CIA)

Table 2.8: Correlation of Supply Shocks 1990-1997

	Por	Lat	Cyp	Gre	Spa	Pol	Slv	UK	Cze	Fra	Bel	Den	Ger	Net	Aus	Swe	Bul	Ita	Ire	Fin	Rom	Hun	Slk	Lit	Lux	Tur
Por		2	.	1	1	1	1	.	..	.	2	1	1	2	1	.	.	1	2	1	1	**2**	1	1	1	.
Lat	2		1	2	1	.	2	1	.	1	2	2	1	.	..	.	.	.	1	.	1	2	2	2	.	..
Cyp	.	1	L	2	1	2	2	1	1	2	**2**	.	2	.	..	.	.	1	.	1	1	.	.	1	1	..
Gre	1	2	2		.	**2**	2	2	2	1	1	...	..	..	.	..	.	..	1	.	.	.	1	1	1	1
Spa	1	1	1	.		1	2	2	.	**2**	1	1	**2**	2	1	1	1	.	2	.	..	1	.	.	..	.
Pol	1	.	2	**2**	1		**2**	1	2	**2**	*	1	1	.	.	..	1	1	..	1	1	1	1	1	1	.
Slv	1	2	2	2	2	**2**		2	2	**2**	*	2	1	.	.	.	1	.	.	..	2	1	..	1	.	.
UK	.	1	1	2	2	1	2		1	2	1	1	..	2	1	1	.	...	1	.	.	.	**2**	1	1	.
Cze	..	.	1	2	.	2	2	1		2	2	2	1	1	.	.	1	...	.	.	1	..	2	...	1	1
Fra	.	1	2	1	**2**	**2**	**2**	2	2	C	**2**	1	2	1	1	.	.	.	.	1	.	.	.	1	1	..
Bel	2	2	**2**	1	1	**2**	**2**	1	2	**2**		**2**	2	1	1	1	1	1	1	1	1	**2**	1	**2**	1	1
Den	1	2	.	...	1	1	2	1	2	1	**2**		1	1	2	1	1	1	.	**2**	.	2	2	2	...	.
Ger	1	1	2	..	**2**	1	1	..	1	2	2	1		2	**2**	2	1	**2**	2	1	1	.	..	..	1	.
Net	2	.	.	..	2	.	.	2	1	1	1	1	2		**2**	1	1	.	1	.	1	.	1	1	2	.
Aus	1	..	..	.	1	.	.	1	.	1	1	2	**2**	**2**		**2**	1	2	2	2	1	.	1	..	.	1
Swe	.	.	.	..	1	..	.	1	.	.	1	1	2	1	**2**		1	1	1	**2**	2	1	1	.	..	.
Bul	.	.	.	.	1	1	1	.	1	.	1	1	1	1	1	1		2	1	2	.	1	1	1	..	2
Ita	1	.	1	..	.	1	.	...	...	.	1	1	**2**	.	2	1	2		1	1	2	1	...	1	.	..
Ire	2	1	.	1	2	..	.	1	.	.	1	.	2	1	2	1	1	1		.	1	1	.	1	1	2
Fin	1	.	1	.	.	1	..	.	.	1	1	**2**	1	.	2	**2**	2	1	.		1	1	**2**	1	..	.
Rom	1	1	1	.	..	1	2	.	1	.	1	.	1	1	1	2	.	2	1	1		1	.	2	1	.
Hun	**2**	2	.	.	1	1	1	.	..	.	**2**	2	.	.	.	1	1	1	1	1	1	R	1	1	...	...
Slk	1	2	.	1	.	1	..	**2**	2	.	1	2	..	1	1	1	1	...	.	**2**	.	1		.	.	..
Lit	1	2	1	1	.	1	1	1	...	1	**2**	2	..	1	..	.	1	1	1	1	2	1	.		.	1
Lux	1	.	1	1	..	1	.	1	1	1	1	...	1	2	.	..	..	.	1	..	1	...	.	.		.
Tur	.	..	..	1	.	.	.	.	1	..	1	.	.	.	1	.	2	..	2	.	.	...	..	1	.	

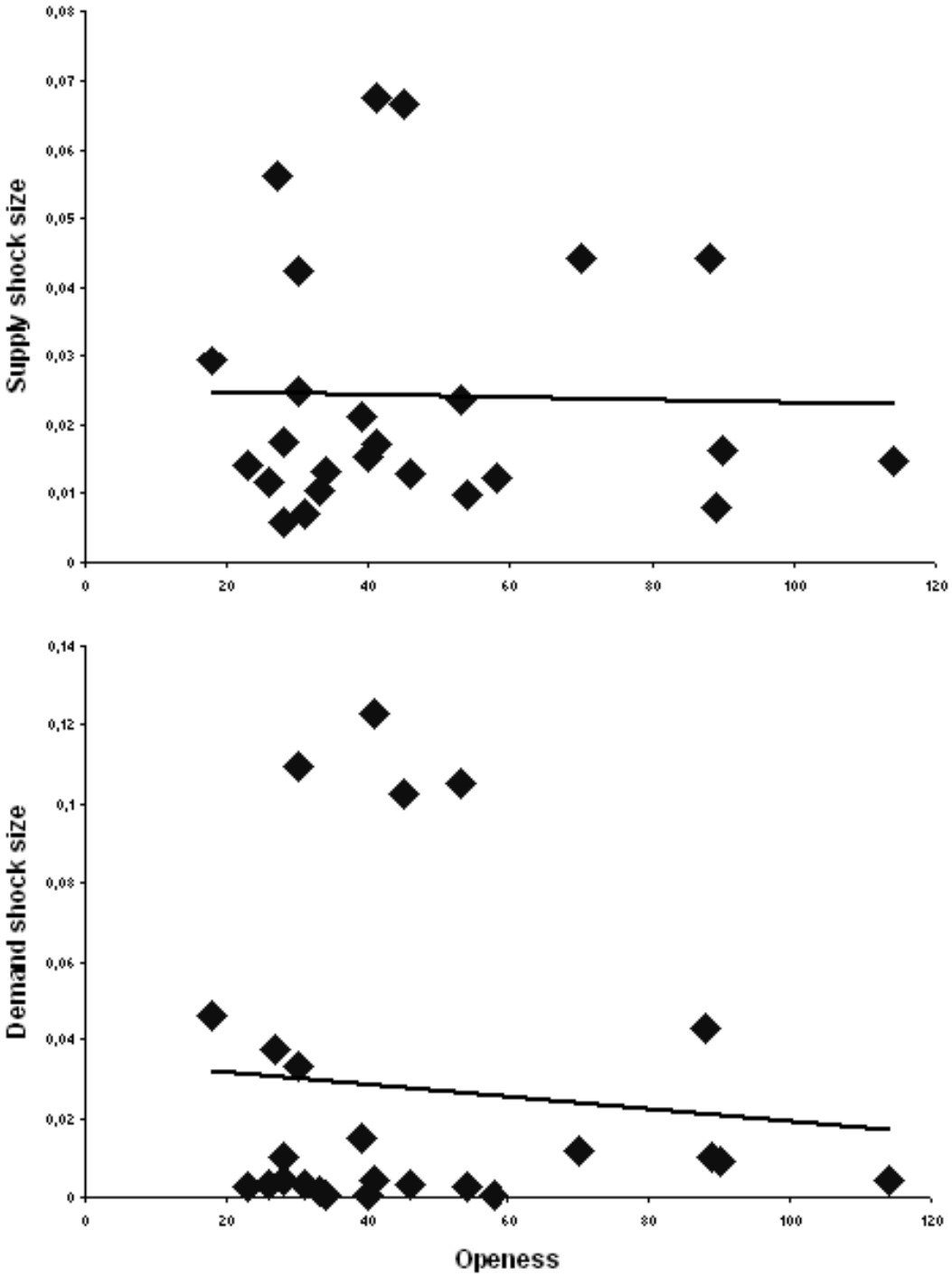
Source: IMF and WIIW; ...,..,.,1,2,3 denote correlations between $-0.36, -0.24, -0.12, 0, 0.16, 0.32, 0.48, 1$;
Significant positive correlation is highlighted by boldfaced numbers.; (C),(L),(R) denote country groupings

Table 2.9: Correlation of Demand Shocks 1990-1997

	Spa	UK	Fra	Den	Ger	Net	Aus	Bel	Gre	Por	Tur	Cyp	Swe	Fin	Pol	Slo	Lit	Ire	Slo	Lux	Lat	Bul	Ita	Hun	Rom	Cze
Spa		**3**	**2**	1	**2**	1	1	2	**2**	**2**	2	**2**	**3**	2	2	1	1	.	1	2	1	.	...	...	.	..
UK	**3**		**3**	2	**3**	2	2	1	**3**	2	1	2	**3**	2	2	2	1	1	1	1	.	..	..	.	.	.
Frae	**2**	**3**	C	2	**2**	**3**	**2**	**2**	**2**	2	**2**	2	**2**	2	1	1	1	1	.	.	.	.	..	.	..	.
Den	1	2	2		2	1	1	**2**	1	1	1	.	**2**	1	.	.	.	1	.	1	.	.	.	.	1	1
Ger	**2**	**3**	**2**	2		**2**	1	1	**2**	**2**	1	2	**2**	1	1	1	.	.	1	2	.	.	..	..	.	.
Net	1	2	**3**	1	**2**		1	**2**	1	2	.	2	2	1	1	1	1	1	1	.	.	.	.	1	..	.
Aus	1	2	**2**	1	1	1		.	2	1	1	2	.	**3**	2	1	.	1	1	..	.	.	..	1	.	.
Bel	2	1	**2**	**2**	1	**2**	.		.	.	2	1	**2**	.	.	.	1	1	1	1	.	.	.	.	1	.
Gre	**2**	**3**	**2**	1	**2**	1	2	.		**2**	**2**	**3**	2	2	1	2	2	1	2	1	1	.	..	.	1	..
Por	**2**	2	2	1	**2**	2	1	.	**2**		2	1	2	1	**2**	2	1	.	.	1	1	.	..	..	.	..
Tur	2	1	**2**	1	1	.	1	2	**2**	2		1	1	1	.	.	.	1	.	1	1	1	.	..	1	.
Cyp	**2**	2	2	.	2	2	2	1	**3**	1	1		1	1	1	1	1	.	.	.	.	.	.	1	1	.
Swe	**3**	**3**	**2**	**2**	**2**	2	.	**2**	2	2	1	1		2	2	1	.	1	.	1	.	.	..	..	..	..
Fin	2	2	2	1	1	1	**3**	.	2	1	1	1	2	E	1	1	.	1	.	..	1	1	..	1	..	.
Pol	2	2	1	.	1	1	2	.	1	**2**	.	1	2	1		**3**	1	.	1	.	1	.	..	1	1	.
Slk	1	2	1	.	1	1	1	.	2	2	.	1	1	1	**3**		1	1	.	.	1	1	..	1	1	.
Lit	1	1	1	.	.	1	.	1	2	1	.	1	.	.	1	1		.	1	.	.	.	.	1	.	2
Ire	.	1	1	1	.	1	1	1	1	.	1	.	1	1	.	1	.		1	1	.	1	.	1	.	.
Slv	1	1	.	.	1	1	1	1	2	.	.	.	.	.	1	.	1	1		.	1	.	1	.	1	..
Lux	2	1	.	1	2	.	..	1	1	1	1	.	1	..	.	.	.	1	.		.	..	.	..	1	.
Lat	1	.	.	.	.	.	.	.	1	1	1	.	.	1	1	1	.	.	1	.	P	2	1	.	2	.
Bul	.	..	.	.	.	.	.	.	.	.	1	.	.	1	.	1	.	1	.	..	2		1	.	.	.
Ita	...	..	..	.	..	.	..	.	..	..	.	.	..	..	..	..	.	.	1	.	1	1		1	1	2
Hun	...	.	.	.	..	1	1	.	.	..	..	1	..	1	1	1	1	1	.	..	.	.	1		1	2
Rom	.	.	..	1	.	..	.	1	1	.	1	1	..	..	1	1	.	.	1	1	2	.	1	1		.
Cze	..	.	.	1	.	.	.	.	..	..	.	.	..	.	.	.	2	.	..	.	.	.	2	2	.	

Source: IMF and WIIW; ...,..,.,1,2,3 denote correlations between $-0.60, -0.40, -0.20, 0, 0.18, 0.37, 0.56$;
Significant positive correlation is highlighted by boldfaced numbers.; (C),(E),(P) denote country groupings

Figure 2.2: Shock Size and Openness 1990-1997



Source: IMF, WDI

Table 2.10: Shock Size 1990-1997

Supply				Demand			
Country	Group			Country	Group		
United Kgd.	EU+4	LE	0.0058	Finland	EU11	RE	0.0004
France	EU11	CC	0.0070	Sweden	EU+4	RE	0.0005
Netherlands	EU11	CC	0.0081	Portugal	EU11		0.0005
Austria	EU11	CC	0.0097	Germany	EU11	CC	0.0013
Germany	EU11	CC	0.0104	Belgium	EU11	CC	0.0014
Spain	EU11	LE	0.0117	Italy	EU11	RP	0.0026
Portugal	EU11		0.0123	Austria	EU11	CC	0.0029
Denmark	EU+4	CC	0.0128	France	EU11	CC	0.0031
Finland	EU11	RE	0.0130	Spain	EU11	LE	0.0032
Belgium	EU11	CC	0.0134	Denmark	EU+4	CC	0.0033
Italy	EU11	RP	0.0140	United Kgd.	EU+4	LE	0.0041
Sweden	EU+4	RE	0.0152	Cyprus	CEE+8		0.0042
Luxembourg	EU11		0.0160	Luxembourg	EU11		0.0069
Ireland	EU11		0.0163	Ireland	EU11		0.0092
Cyprus	CEE+8		0.0171	Greece	EU+4	LE	0.0100
Greece	EU+4	LE	0.0174	Netherlands	EU11	CC	0.0102
Slovenia	CEE5	RP	0.0210	Czech Rep.	CEE5		0.0116
Hungary	CEE5	RP	0.0234	Slovenia	CEE5	RP	0.0150
Poland	CEE5		0.0248	Poland	CEE5		0.0330
Turkey	CEE+8		0.0294	Lithuania	CEE+8		0.0372
Romania	CEE+8	RP	0.0423	Slovak Rep.	CEE+8		0.0429
Czech Rep.	CEE5		0.0442	Turkey	CEE+8		0.0458
Slovak Rep.	CEE+8		0.0444	Latvia	CEE+8		0.1027
Lithuania	CEE+8		0.0562	Hungary	CEE5	RP	0.1052
Latvia	CEE+8		0.0666	Romania	CEE+8	RP	0.1094
Bulgaria	CEE+8	RP	0.0676	Bulgaria	CEE+8	RP	0.1227

Source: IMF, WIIW

Table 2.11: Shock Size Summary 1990-1997

Supply Group			Demand Group		
Average					
	CC	0.0102		RE	0.0005
EU11		0.0120		CC	0.0037
EU+4		0.0128	EU11		0.0038
	LE	0.0130	EU+4		0.0045
	RE	0.0141		LE	0.0054
All		0.0238	All		0.0265
CEE5		0.0284	CEE5		0.0412
	RP	0.0337	CEE+8		0.0664
CEE+8		0.0462		RP	0.0710
Coefficient of Variation					
	RE	0.0766		RE	0.0897
	CC	0.2236		LE	0.5044
EU11		0.2369	CEE+8		0.6242
CEE5		0.3262		RP	0.7216
EU+4		0.3399	EU+4		0.7748
	LE	0.3633		CC	0.8096
CEE+8		0.3767	EU11		0.8545
	RP	0.5759	CEE5		0.9177
All		0.7473	All		1.4328

Source: IMF, WIIW

Table 2.12: Speed of Adjustment 1990-1997

Supply				Demand			
Country	Group			Country	Group		
Belgium	EU11	CC	0.20	Sweden	EU+4	RE	0.21
Cyprus	CEE+8		0.48	Portugal	EU11		0.27
Slovenia	CEE5	RP	0.53	Latvia	CEE+8		0.42
Poland	CEE5		0.54	Germany	EU11	CC	0.52
Austria	EU11	CC	0.64	Greece	EU+4	LE	0.65
Turkey	CEE+8		0.66	Luxembourg	EU11		0.68
Luxembourg	EU11		0.70	Ireland	EU11		0.69
Spain	EU11	LE	0.72	Finland	EU11	RE	0.69
Netherlands	EU11	CC	0.72	Slovenia	CEE5	RP	0.69
Ireland	EU11		0.72	Turkey	CEE+8		0.70
Portugal	EU11		0.76	Austria	EU11	CC	0.74
Greece	EU+4	LE	0.78	Cyprus	CEE+8		0.80
Germany	EU11	CC	0.78	Czech Rep.	CEE5		0.86
Latvia	CEE+8		0.80	Spain	EU11	LE	0.87
Sweden	EU+4	RE	0.84	Hungary	CEE5	RP	0.88
Hungary	CEE5	RP	0.85	Bulgaria	CEE+8	RP	0.90
France	EU11	CC	0.86	Italy	EU11	RP	0.91
Czech Rep.	CEE5		0.86	Slovak Rep.	CEE+8		1.02
United Kgd.	EU+4	LE	0.89	Lithuania	CEE+8		1.03
Italy	EU11	RP	0.92	United Kgd.	EU+4	LE	1.09
Denmark	EU+4	CC	0.94	Poland	CEE5		1.11
Romania	CEE+8	RP	0.94	Denmark	EU+4	CC	1.20
Finland	EU11	RE	0.94	France	EU11	CC	1.30
Lithuania	CEE+8		1.05	Belgium	EU11	CC	1.58
Bulgaria	CEE+8	RP	1.06	Netherlands	EU11	CC	2.53
Slovak Rep.	CEE+8		1.30	Romania	CEE+8	RP	3.31

Source: IMF, WIIW

Table 2.13: Speed of Adjustment Summary 1990-1997

Supply Group			Demand Group		
Average					
	CC	0.6891		RE	0.4508
CEE5		0.6933	EU+4		0.7881
	LE	0.7166		LE	0.8538
EU11		0.7225	CEE5		0.8853
All		0.7871	EU11		0.9792
	RP	0.8592	All		0.9862
EU+4		0.8632	CEE+8		1.1680
	RE	0.8920		CC	1.3103
CEE+8		0.8986		RP	1.3388
Coefficient of Variation					
	RE	0.0566	EU11		0.1679
EU11		0.0683		LE	0.1835
	RP	0.2098	CEE5		0.4952
	LE	0.2121		CC	0.4953
EU+4		0.2331		RE	0.5325
CEE5		0.2601	EU+4		0.6095
All		0.2682	All		0.6498
CEE+8		0.2853		RP	0.7394
	CC	0.3452	CEE+8		0.7679

Source: IMF, WIIW

Table 2.14: Dominant Collective Bargaining Structure Over the Past 10 Years

Czech Rep.	C	Bulgaria	N/S, C	Austria	N/S
Estonia	C	Finland	N/S, C	Belgium	N/S
Hungary	C	Slovenia	N/S	France	N/S
Ireland	C	Cyprus	N/S	Germany	N/S
Poland	C	Denmark	N/S	Italy	N/S
Slovakia	C	Sweden	N/S	Netherlands	N/S
United Kingdom	C	Greece	N/S	Portugal	N/S
				Spain	N/S

Source: World Labour Report 1997-98 (ILO); N/S = National/sectoral level; C = Company/plant level

Table 2.15: Progress of the Transformation Process 1997

	Political stability	Currency	Privatization	Legal system	Banking sector	Economic development	Infrastructure	Average
Poland	8	8	8	8	9	9	8	8.29
Hungary	8	7	8	9	9	8	8	8.14
Czech Republic	7	8	8	9	7	8	9	8.00
Estonia	7	9	9	8	9	8	6	8.00
Latvia	7	8	8	6	7	7	7	7.14
Slovenia	7	7	4	8	8	8	8	7.14
Lithuania	6	7	6	5	6	7	6	6.14
Slovak Republic	4	7	5	6	7	7	7	6.14
Romania	5	5	5	6	4	6	5	5.14
Bulgaria	5	5	5	6	4	5	5	5.00

Source: Rödl & Partner Research, (FAZ GmbH Informationsdienste 1998)

Table 2.16: SURPLUS (% OF GDP)

Group	Country	Average	1993	1994	1995	1996	1997	1998
EU11	Luxembourg	2.4		0.5	2.3	4.6	2.0	
EU11	Ireland	0.1	-0.8	-0.9	-0.6	0.2	0.5	2.0
CEE5	Estonia	0.1	-2.1	1.4	-0.6	-0.8	2.5	-0.1
CEE5	Czech Rep.	-0.2	0.1	0.9	0.5	-0.1	-1.0	-1.6
CEE5	Slovenia	-0.4	0.4	-0.3	-0.3	0.1	-1.5	-0.6
EU11	Finland	-0.9	-1.3	-1.1	-0.9	-0.6	-0.2	
EU11	Netherlands	-1.4	-1.4	0.2	-3.6	-1.4	-1.5	-0.7
EU11	Germany	-1.7	-2.4	-1.3	-1.8	-2.1	-1.4	-0.9
CEE5	Poland	-1.7		-2.1	-1.9	-2.0	-1.3	-1.0
CEE+8	Latvia	-2.0	-3.5	-4.0	-3.7	-1.6	0.7	0.1
EU+4	Denmark	-2.1	-2.2	-2.3	-1.9			
CEE+8	Romania	-2.8	-0.5	-2.5	-3.0	-4.0	-3.9	
CEE+8	Cyprus	-3.2	-2.4	-1.4	-1.0	-3.4	-5.3	-5.6
CEE+8	Lithuania	-3.6	-6.0	-4.7	-4.8	-3.6	-1.9	-0.4
EU11	Belgium	-3.6	-6.1	-4.2	-3.2	-2.5	-2.0	
EU+4	United Kgd.	-3.9	-7.3	-5.9	-5.5	-3.6	-2.0	0.6
EU11	Portugal	-4.2	-6.9	-4.8	-5.0	-2.3	-2.1	
CEE+8	Slovak Rep.	-4.3	-6.2	-5.2	-1.6	-4.2	-3.7	-4.9
EU11	Spain	-4.4	-6.1	-6.4	-4.8	-5.4	-2.5	-1.3
EU11	Austria	-4.5	-5.0	-5.8	-5.0	-4.1	-2.7	
EU11	France	-5.2	-5.6	-5.5	-6.5	-5.2	-3.5	
CEE+8	Malta	-5.4	-2.9	-3.7	-2.7	-7.7	-9.8	
CEE5	Hungary	-5.5	-5.7	-7.1	-6.4	-3.1	-4.5	-6.2
CEE+8	Turkey	-5.7	-6.5	-3.9	-4.1	-8.4		
CEE+8	Bulgaria	-6.0	-12.1	-4.7	-5.2	-19.0	2.1	2.8
EU+4	Sweden	-6.2	-16.0	-8.4	-8.9	-3.3	-0.9	0.4
EU11	Italy	-6.2	-10.1	-9.3	-6.9	-7.2	-1.6	-2.4
EU+4	Greece	-13.6	-11.5	-21.1	-11.9	-9.7		
Average		-3.4						
CEE+8		-1.5	-1.8	-1.4	-1.7	-1.2	-1.1	-1.9
EU11		-2.7	-4.6	-3.5	-3.3	-2.4	-1.4	-0.7
CEE5		-4.1	-5.0	-3.8	-3.3	-6.5	-3.1	-1.6
EU+4		-6.5	-9.2	-9.4	-7.1	-5.5	-1.5	0.5

Source: IMF

Table 2.17: SEIGNORAGE (% OF GDP)

Group	Country	Average	1989	1990	1991	1992	1993	1994	1995
EU11	France	-0.1	0.0	-0.2	-0.2	-0.7	-0.2	0.0	0.3
EU11	Belgium	0.1	0.2	-0.2	0.1	0.0	0.2	-0.4	0.5
EU11	Luxembourg	0.1	0.0	0.2	0.0	0.0	0.5	0.0	-0.1
EU+4	United Kgd.	0.2	0.3	0.0	0.0	0.2	0.3	0.2	0.3
EU11	Spain	0.3	3.9	-6.0	2.5	0.0	0.1	1.2	0.5
EU11	Germany	0.3			0.5	1.3	-0.1	-0.2	0.2
EU11	Netherlands	0.4	1.3	0.3	-1.2	2.1	0.7	0.6	-1.2
EU11	Austria	0.5	1.2	0.0	0.2	0.9	0.7	0.4	-0.1
EU11	Ireland	0.5	0.4	0.9	-0.8	-0.6	1.3	0.3	1.8
EU+4	Denmark	0.5	-0.2	1.0	-0.2	-0.6	2.8	0.0	0.9
EU11	Italy	0.6	1.8	1.2	0.7	0.6	-1.3		
EU11	Finland	0.8	1.5	-1.3	0.7	0.4	0.0	3.8	0.2
EU+4	Sweden	0.9	0.7	0.5	-0.5	1.5	3.7	2.4	-1.8
CEE+8	Latvia	1.2						2.1	0.2
EU11	Portugal	1.2	11.1	1.6	4.1	2.1	1.8	-12.3	-0.2
CEE5	Slovenia	1.4				2.1	1.0	1.6	0.9
EU+4	Greece	2.1	1.4	3.3	1.7	1.5	1.2	5.3	0.6
CEE+8	Lithuania	3.0							3.0
CEE+8	Slovak Rep.	3.7						2.1	5.4
CEE5	Estonia	5.3				8.6	9.0	1.5	2.0
CEE5	Hungary	5.4	3.4	8.0	11.0	3.0	3.7	3.6	
CEE+8	Romania	5.8	8.9	8.4	3.3	7.7	5.8	3.6	3.0
CEE5	Poland	6.3	22.9	9.4	3.0	3.4	0.7	1.7	3.1
CEE5	Czech Rep.	11.3						5.7	16.9
Average		2.2							
EU11		0.4	2.1	-0.4	0.6	0.6	0.3	-0.7	0.2
EU+4		0.9	0.6	1.2	0.3	0.6	2.0	2.0	0.0
CEE+8		3.4	8.9	8.4	3.3	7.7	5.8	2.6	2.9
CEE5		5.9	13.1	8.7	7.0	4.3	3.6	2.8	5.7

Source: World Development Indicators (World Bank)

Table 2.18: National Unemployment Variation 1998

	Highest	Mean	Lowest	Highest/Lowest	Coefficient of variation
EU11+4+CEE5	20.8	9.5	3.6	5.8	1.13
CEE+8	13.7	7.5	3.4	4.0	0.47
EU11	20.8	10.7	3.6	5.8	0.47
EU11+4	20.8	9.9	3.6	5.8	0.45
EU11+CEE5	20.8	9.9	3.6	5.8	0.44
EU+CEE5	20.8	8.9	3.4	6.1	0.44
CEE5	11.5	8.4	4.8	2.4	0.31
USA	8.8	5.6	2.5	3.5	0.21

Source: IMF

Table 2.19: Regional Unemployment Variation 1998

	Highest	National	Lowest	Highest/Lowest
EU11+4	31.8	10.9	3.9	8.2
USA	8.8	5.6	2.5	3.5
Italy	25.9	11.9	6.0	4.3
Portugal	11.4	7.3	3.9	2.9
Denmark	12.3	7.1	4.5	2.7
Germany	11.2	6.6	5.5	2.0
Belgium	12.9	9.9	6.9	1.9
France	15.3	11.5	8.7	1.8
Spain	31.8	22.9	18.5	1.7
Sweden	12.0	9.2	7.7	1.6
United Kgd.	11.0	8.8	6.7	1.6
Greece	11.0	9.1	7.4	1.5
Finland	21.2	17.2	15.2	1.4
Ireland	15.8	14.3	13.5	1.2
Netherlands	8.9	7.3	6.9	1.2

Source: European Commission Services

Table 2.20: Factor Payments and Receipts 1997 (% of GDP)

Group	Country	Sum	Receipts	Payments
EU11	Belgium	43.4	23.0	20.4
EU+4	Denmark	42.2	20.0	22.2
EU11	Ireland	32.5	9.8	22.7
EU+4	United Kgd.	26.3	13.9	12.4
CEE+8	Malta	19.7	10.7	9.0
EU11	Netherlands	17.3	9.3	8.0
EU+4	Sweden	15.2	6.2	8.9
EU11	Austria	9.8	4.8	5.0
CEE5	Hungary	9.2	3.0	6.1
EU11	Italy	9.0	4.0	5.0
EU11	Finland	8.6	3.1	5.5
CEE5	Estonia	8.0	2.4	5.6
EU11	Portugal	7.9	3.8	4.1
CEE+8	Bulgaria	7.7	2.1	5.6
EU11	Germany	7.4	3.7	3.8
EU11	France	7.4	3.8	3.6
CEE5	Czech Rep.	6.9	2.7	4.2
EU11	Spain	6.1	2.5	3.7
CEE+8	Latvia	5.4	3.2	2.2
CEE+8	Cyprus	4.6	1.7	2.9
CEE+8	Slovak Rep.	3.9	1.6	2.3
CEE5	Slovenia	3.9	2.3	1.6
CEE+8	Lithuania	3.7	0.8	2.9
EU+4	Greece	3.7	0.9	2.7
CEE+8	Turkey	3.6	1.0	2.6
CEE5	Poland	3.0	1.1	1.9
CEE+8	Romania	2.1	0.6	1.5
Average		11.8	5.3	6.5
EU+4		21.8	10.3	11.6
EU11		15.0	6.8	8.2
CEE5		6.3	2.7	3.6
CEE+8		6.2	2.3	3.9

Source: World Development Indicators (World Bank)

Table 2.21: Inflation Rate p.a. 1992−2000

Bulgaria	87.52	Cyprus	3.37
Romania	85.34	Malta	3.31
Turkey	74.46	Italy	3.27
Lithuania	57.91	United Kgd.	2.82
Estonia	39.83	Ireland	2.70
Latvia	37.68	Netherlands	2.40
Poland	20.25	Austria	2.26
Hungary	17.57	Denmark	2.22
Slovenia	17.07	Luxembourg	2.15
Slovak Rep.	10.53	Germany	2.13
Czech Rep.	8.98	Belgium	2.12
Greece	7.20	Finland	1.79
Portugal	3.93	France	1.58
Spain	3.47	Sweden	1.54

Source: IMF

Table 2.22: Export plus Import (% of GDP)

Malta	125	Sweden	40
Belg./Lux.	114	Slovenia	39
Ireland	90	Finland	34
Netherlands	89	Germany	33
Slovak Rep.	88	France	31
Estonia	72	Romania	30
Czech Rep.	70	Poland	30
Portugal	58	Greece	28
Austria	54	United Kingdom	28
Hungary	53	Lithuania	27
Denmark	46	Spain	26
Latvia	45	Italy	23
Cyprus	41	Turkey	18
Bulgaria	41		

Source: COMEXT (Eurostat), WIIW, 1995

Table 2.23: Results

		Real ex. rate changes	Trade specialization	FDI	Supply shock corr.	Demand shock corr.	Supply & demand Group	CC or RE or LE	Wage bargaining	Transform. process	Deficit	Seignorage	Openness	Suitability index
Weights ex real exchange rate changes			0.2	0.05				0.2	0.1	0.15	0.05	0.05	0.2	
Weights		0.1	0.18	0.045				0.18	0.09	0.135	0.045	0.045	0.18	
EU+4	Denmark	+++	+	+	CR	CE	CC	+	-	+	+	+	+	1.11
EU11	Netherlands	+++	+	-	CR	CE	CC	+	-	+	+	+	+	1.07
EU11	Austria	+++	+	+	CR	CE	CC	+	-	+	-	+	+	1.07
EU11	Belgium	+++	+	-	CLR	CE	CC	+	-	+	-	+	+	1.02
EU11	Germany	+++	+	+	CR	CE	CC	+	-	+	+	+	-	0.93
EU11	Luxembourg	+++	+	-	-	-	-	-	-	+	+	+	+	0.89
EU11	France	++	+	+	CL	CE	CC	+	-	+	-	+	-	0.79
EU11	Finland	++	-	+	R	E	RE	+	+	+	+	+	-	0.74
EU11	Spain	+	+	+	L	E	LE	+	-	+	-	+	-	0.69
EU+4	United Kgd.	-	+	+	L	E	LE	+	+	+	-	+	-	0.68
EU+4	Sweden	+	-	-	R	E	RE	+	-	+	-	+	+	0.64
EU11	Ireland	+	-	-	R	-	-	-	+	+	+	+	+	0.60
CEE+8	Cyprus	-	-	+	L	E	LE	+	-	+	-		+	0.54
EU11	Portugal	+	-	+	-	E	-	-	-	+	-	+	+	0.51
CEE5	Slovenia	-	+	+	L	-	RP	-	-	-	+	+	+	0.50
EU+4	Greece	+	-	+	L	E	LE	+	-	+	-	-	-	0.46
CEE5	Estonia	-	-	-					+	+	+	-	+	0.45
CEE5	Czech Rep.	-	+	-	L	P	LP	-	+	+	+	-	-	0.45
CEE5	Poland	-	+	-	L	-	-	-	+	+	+	-	-	0.45
CEE+8	Malta	+	-	-					-	+	-		+	0.42
EU11	Italy	-	+	+	R	P	RP	-	-	+	-	+	-	0.41
CEE5	Hungary	-	-	-	R	P	RP	-	+	+	-	-	+	0.41
CEE+8	Bulgaria	-	-	+	R	P	RP	-	+	-	-		+	0.32
CEE+8	Slovak Rep.	-	+	+	-	-	-	-	+	-	-	-	-	0.32
CEE+8	Turkey	+	-	+	-	E	-	-	-	+	-		-	0.28
CEE+8	Latvia	-	-	-	-	P	-	-	-	-	+	+	+	0.27
CEE+8	Romania	-	-	+	R	P	RP	-	-	-	+	-	-	0.09
CEE+8	Lithuania	-	-	+	-	-	-	-	-	-	-	-	-	0.05

2.6.3 Data sources

Source	Data
IMF (International Financial Statistics 11/2000)	Nominal exchange rates Price indices Deficits
IMF (International Financial Statistics 11/2000) and Vienna Institute for International Economic Studies - WIIW	Industrial production index GDP
World Fact Book (CIA) at http://www.cia.gov/cia/publications /factbook/index.html	Sectoral GDP shares GDP_{PPP}
Bureau of Labor Statistics at http://www.bls.gov/sahome.html	Regional price indices
FAZ GmbH Informationsdienste (1998)	Transformation process index
World development indicators (World Bank)	Seignorage, FDI Factor payments plus receipts
COMTRADE (United Nations) at http://www.intracen.org COMEXT (Eurostat)	Im/export flows
World Labour Report (ILO)	Collective bargaining structure

The price indices for the following US regions are used: New York-Northern New Jersey-Long Island, Philadelphia-Wilmington-Atlantic City, Detroit-Ann Arbor-Flint, Cleveland-Akron, Boston-Brockton-Nashua, Chicago-Gary-Kenosha, Dallas-Fort Worth, Houston-Galveston-Brazoria, Miami-Fort Lauderdale, Los Angeles-Riverside-Orange County, San Francisco-Oakland-San Jose.

Chapter 3

Does the Nominal Exchange Rate Facilitate Real Exchange Rate Adjustment in Floating Exchange Rate Regimes? A SVAR Decomposition of German Exchange Rates in 1973-1998

3.1 Introduction

The relative merits of flexible and fixed exchange rate regimes are closely related to the ability of the nominal exchange rate to absorb shocks between national economies. In the definition of the real exchange rate

$$\begin{aligned} Q &= \frac{S * P^*}{P} \\ \hat{Q} &= \hat{S} + \hat{P^*} - \hat{P} \end{aligned}$$

we see that in case of a fixed nominal exchange rate, i.e. $\hat{S} = 0$, changes in the real exchange rate $\hat{Q} \neq 0$ have to be accomplished by differences in the relative inflation rates, i.e. $\hat{P^*} - \hat{P} \neq 0$. The nominal exchange rate potentially facilitates adjustments in the real exchange rate. Especially, a changing real exchange rate implies that both countries cannot achieve a zero, or any other optimal inflation rate, simultaneously. If relative costs of adjustment through prices rather than nominal exchange rates are higher it is preferable that the exchange rates adjust in case of shocks. Nominal exchange rates are wholesale prices reacting instantaneously to news in contrast to sticky prices so that periods of misalignment with resulting misallocation can be expected to be shorter for

the appropriate real exchange rate changes. The theory of optimum currency areas argues in the same vein and identifies macroeconomic adjustment costs of currency areas where the nominal exchange rate is not available as an adjustment instrument.
The model of Obstfeld (1985) is used to derive long run restrictions in order to identify structural shocks by the method of Blanchard and Quah (1989). It is a stock flow consistent linear stochastic rational expectations two country model reproducing Mundell-Fleming-Dornbusch effects. Clarida and Gali (1994) and Chadha and Prasad (1997) use the same model for a real exchange rate decomposition. The first paper takes the USA and the second paper takes Japan as the home country whereas here the estimation is done for Germany.
The hypothesis analyzed is formulated in terms of impulse response functions.

3.2 Economic Model

The estimation and discussion are based upon an open economy two country model.

$$
\begin{aligned}
y_t^d &= \eta q_t - \sigma[i_t - E_t(p_{t+1} - p_t)] + d_t && (3.1)\\
m_t - p_t &= y_t^d - \lambda i_t && (3.2)\\
p_t &= E_{t-1}\tilde{p}_t + \theta(\tilde{p}_t - E_{t-1}\tilde{p}_t && (3.3)\\
&= \theta\tilde{p}_t + (1-\theta)E_{t-1}\tilde{p}_t && (3.4)\\
i_t &= E_t(s_{t+1} - s_t) && (3.5)\\
q_t &\equiv s_t - p_t && (3.6)
\end{aligned}
$$

Expectations $E_j(\tilde{x}_t)$ are conditioned on information I_j in period j. The variables y, m, p, i refer to relative output, money, prices and interest rates, respectively. All variables, except interest rates, are used in natural logarithms and represent relative values as differences between home (h) and foreign (f) values. Definition (3.6) for the real exchange rate is rather unusual because $p_t = p_t^h - p_t^f$. Foreign nominal interest rates and foreign prices are set to zero $p_t^f = i_t^f = 0$ to facilitate the algebra. The model explains the endogenous variables relative aggregate demand y_t^d, relative price p_t and the real exchange rate q_t. The nominal exchange rate s_t is endogenous, too, since it is the sum of the endogenous real exchange rate and a relative price.
Equation (3.1) is an open economy IS relationship and relates an increasing relative aggregate demand y_t^d to an increasing (=depreciating) real exchange rate q_t since home goods became more competitive; to decreases of the real interest rate difference, for example due to investment demand effects of interest rate changes; and finally to an increase in the residual relative absorption.

Equation (3.2) represents the LM money market equilibrium, where the relative real money demand increases with the relative output y_t at an assumed elasticity of one and decreases with the relative nominal interest rate i_t.
Equations (3.3) and (3.4) describe the price adjustment process in a sticky price environment. The flexible price equilibrium values will be labelled by a tilde. An instantaneous price adjustment is characterized by $\theta = 1$ and a sluggish price adjustment by $0 < \theta < 1$. The market clearing price that was expected in the previous period $E_{t-1}\tilde{p}_t$ (the predetermined part of the relative price) adjusts only partially to new information $\tilde{p}_t - E_{t-1}\tilde{p}_t$. Equation (3.4) expresses the relative price as a weighted average of the predetermined and the equilibrium relative price levels.
The hypothetical long run solution, where prices are flexible, is used as an equilibrium concept instead of a steady state. The disequilibrium due to sticky prices is characterized by a gap between the sticky price solution and the long-run equilibrium.
The uncovered interest parity in equation (3.5) relates the nominal interest rate difference to the expected nominal exchange rate changes. Risk premia are assumed to be constant.
Shocks are specified through the processes of the exogenous variables.

$$\begin{aligned} y_t^s &= y_{t-1}^s + z_t \\ m_t &= m_{t-1} + v_t \\ d_t &= d_{t-1} + \delta_t - \gamma\delta_{t-1} \end{aligned} \tag{3.7}$$

These are relative real supply shocks z_t (AS), representing shifts in the aggregate supply curve, relative monetary shocks v_t (LM) representing a relative money demand and supply shifts and relative real demand shocks δ (IS), representing relative domestic absorption shifts in consumption, government expenditure, investment or tastes. They all follow white noise $N(0, \sigma_i^2)$ processes. Since the shocks are modelled in relative terms they represent asymmetric shifts which trigger the need for an exchange rate adjustment. The supply (AS) and monetary (LM) shocks are entirely permanent since they follow pure random walk processes. The demand shocks (IS) on the other hand possess transitory dynamics in addition to the permanent component, since shocks in t are partly reversed in period $t+1$ through $-\gamma\delta_{t-1}$.

3.2.1 Long run solutions

The following expressions are derived in section (3.5.1.1) of the appendix and characterize the long-run solutions for the endogenous variables.

$$\tilde{y}_t = y_t^s \tag{3.8}$$

$$\tilde{q}_t = \frac{y_t - d_t}{\eta} + \frac{\sigma}{\eta(\eta+\sigma)}\gamma\delta_t \tag{3.9}$$

$$\tilde{p}_t = m_t - y_t + \frac{\lambda}{(1+\lambda)(\eta+\sigma)}\gamma\delta_t \tag{3.10}$$

$$\tilde{q}_t + \tilde{p}_t \equiv \tilde{s}_t = m_t + \frac{1-\eta}{\eta}y_t - \frac{d_t}{\eta} + \left(\frac{\sigma}{\eta(\eta+\sigma)} + \frac{\lambda}{(1+\lambda)(\eta+\sigma)}\right)\gamma\delta_t \tag{3.11}$$

Equation (3.8) describes the output when prices adjust to the good market clearing level. The relative output is determined entirely by supply (AS) shocks and can be represented by a vertical long run supply curve. The absence of demand (IS) and monetary (LM) shocks (money neutrality) gives two out of three identification restrictions, distinguishing AS from IS and LM shocks.

Identification 1 *Relative real demand (IS) shocks have no long run effects on the relative real output level.*

Identification 2 *Relative monetary (LM) shocks have no long run effects on the relative real output level.*

Equation (3.9) is derived in section (3.5.1.1.1) of the appendix and corresponds to equation (3.22). This solution shows that monetary (LM) shocks have no long run effects on the real exchange rate, which gives the last out of three identification restrictions, distinguishing IS from LM shocks.

Identification 3 *Relative monetary (LM) shocks have no long run effect on the real exchange rate.*

The rationale of this finding is that monetary LM shocks induce exactly the same effects on the price differential and the nominal exchange rate in the long run, represented by the fact that m_t is contained in the solution for the relative price (3.10). The nominal exchange rate (3.11) cancels out in the real exchange rate definition (3.6) $q_t = s_t - p_t$. Positive relative real supply shocks (AS) result in a real exchange rate depreciation and a permanent demand shock (IS) leads to an appreciation if $1/\eta > \gamma\sigma(\eta(\eta+\sigma))^{-1}$. Furthermore, since γ is positive there is no effect of the temporary component of the IS shock on the real exchange rate. Equation (3.10) is derived in section (3.5.1.1.2) of the appendix and corresponds to equation (3.26). Supply shocks (AS) reduce the relative price as the consequence of a shift of the vertical long run supply curve to the right. Monetary (LM) shocks raise relative price as well as the temporary component of the real

demand shock (IS). Graphically this is due to an upward shift of the aggregate real demand curve along the vertical long run supply curve. There is no effect of the permanent part of the demand shocks (IS) on the relative price.
Equation (3.11) is simply the sum of the solutions for the real exchange rate (3.9) and the relative price (3.10). Monetary shocks (LM) depreciate (raise) the nominal exchange rate in the long run. Most interesting, the real demand and supply shocks (AS, IS) have theoretically ambiguous effects on the nominal exchange rate. This is due to the fact that these shocks have opposite effects on the relative price and the real exchange rate. Therefore the effect of these shocks on the sum $s_t \equiv q_t + p_t$ is indeterminate. The ability of the nominal exchange rate to facilitate the real adjustment in flexible exchange rate regimes is measured empirically by the extent to which the effects of real shocks (AS, IS) on the real exchange rate corresponds to their effect on the nominal exchange rate.

Hypothesis 1 *The nominal exchange rate facilitates the real adjustment in flexible exchange rate regimes, if the real demand and supply shocks (AS, IS) affect the nominal exchange rate in the same way as they affect the real exchange rate.*

Or put it in a qualitative way: the nominal exchange rate facilitates the adjustment, if the reaction of the real exchange rate to demand (IS) and supply (AS) shocks is not entirely carried out by the relative price but partly by nominal exchange rate movements, thus easing the burden of real adjustment for relative prices.
Another interesting feature of the model is that it exhibits a triangular structure since the relative output is affected only by supply shocks, the real exchange rate by supply and demand shocks and the relative price by all of them. This can be used in the specification of the econometric model.

3.2.2 Sticky-price solutions

Sticky prices are represented by θ between zero and one. In this case the output is demand determined. The following equations constitute the short run solutions which are derived in section (3.5.1.2) of the appendix.

$$p_t = m_{t-1} - y_{t-1} + \theta(v_t - z_t + \alpha\gamma\delta_t) \tag{3.12}$$

$$q_t = \frac{y_t - d_t}{\eta} + \frac{\gamma(\sigma+\lambda) - \eta\alpha\gamma\theta(1+\lambda)}{\eta(\eta+\sigma+\lambda)} d_t + \frac{(1-\theta)(1+\lambda)}{\eta+\sigma+\lambda}(v_t - z_t) \tag{3.13}$$

$$s_t = \frac{1-\eta\theta}{\eta} y_t^s - \frac{d_t}{\eta} + \left(\frac{(1+\lambda)(1-\theta)}{\eta+\sigma+\lambda} - (1-\theta)\right)(v_t - z_t) + \theta m_t \tag{3.14}$$

$$\left(\frac{1}{\eta(\eta+\sigma)} + \frac{\theta\lambda}{(1+\lambda)(\eta+\sigma)} - \frac{(1+\lambda)(1-\theta)}{\eta+\sigma+\lambda}\right)\gamma\delta_t \tag{3.15}$$

The solutions are compared with the long run equilibrium to identify over- and undershooting behavior:

$$p_t = \tilde{p}_t - (1-\theta)(v_t - z_t + \alpha\gamma\delta_t) \tag{3.16}$$

$$q_t = \tilde{q}_t + \frac{(1-\theta)(1+\lambda)}{\eta+\sigma+\lambda}(v_t - z_t + \alpha\gamma\delta_t) \tag{3.17}$$

$$s_t = \tilde{s}_t + (1-\eta-\sigma)\frac{1-\theta}{\eta+\sigma+\lambda}(v_t - z_t + \alpha\gamma\delta_t) \tag{3.18}$$

$$y_t^d = y_t^s + \frac{(1-\theta)(1+\lambda)(\eta+\sigma)}{\eta+\sigma+\lambda}(v_t - z_t + \alpha\gamma\delta_t)) \tag{3.19}$$

The difference between the short and the long run equilibrium value of the relative price in equation (3.16), is derived in section 3.5.1.2.1 of the appendix and corresponds to equation (3.28). Expansionary demand (IS) and monetary (LM) shocks increase the relative price, whereas supply (AS) shocks decrease the relative price in equation (3.12). The extent of the disequilibrium (p_t-$\tilde{p}_t$) is proportional to the contemporaneous news $v_t - z_t + \alpha\gamma\delta_t$ and increases with the degree of price sluggishness $(1-\theta)$. Monetary (LM) and demand (IS) shocks push the relative price below and supply (AS) shocks above the long run level in equation (3.16). The disequilibrium exists only one period since no lagged shocks enter the equation.
Equation (3.17), the solution for the real exchange rate, is derived in section 3.5.1.2.2 of the appendix and corresponds to equation (3.31). Monetary (LM) shocks depreciate the real exchange rate due to price rigidities, supply (AS) shocks cause an appreciation and demand shocks have an ambiguous effect in equation (3.13). Monetary (LM) and demand (IS) shocks result in the real exchange rate above and supply (AS) shocks below the long-run level in equation (3.17).
Equation (3.18), the solution for the nominal exchange rate, is again the sum of the short run solutions of the relative price (3.16) and the real exchange rate (3.17). If $1-\eta-\sigma > 0$, monetary (LM) and demand (IS) shocks cause the nominal exchange rate to overshoot and supply (AS) shocks to undershoot the long-run level. In contrast to the Dornbusch model this model produces overshooting also due to real demand shocks.
Equation (3.19), the solution for the relative output, is derived in section 3.5.1.2.3 of the appendix and corresponds to equation (3.32). Monetary (LM) shocks raise the relative demand, therefore the monetary policy can be used for a stabilization policy in flexible exchange rate regimes. The temporary part of demand (IS) shocks ($\gamma\delta_t$) raise the relative output above the long run level, whereas the effect of supply (AS) shocks is reduced by price stickiness.
Section (3.5.2) of the appendix shows why three additional identifying assumptions are needed. They concern the sign of the effect which each shocks has on an other variable. This is because the Blanchard-Quah methodology provides no guidance whether positive or negative shocks are identified. Therefore the following three theoretically derived effects will have to be assumed a priori:

Identification 4 *Supply (AS) shocks raise the relative output on impact.*

Identification 5 *Demand (IS) shocks raise the relative price on impact.*

Identification 6 *Monetary (LM) shocks raise the relative price on impact.*

3.3 Estimation

Since the model is specified in relative terms bilateral figures are used for estimation. The estimated VAR includes three variables:

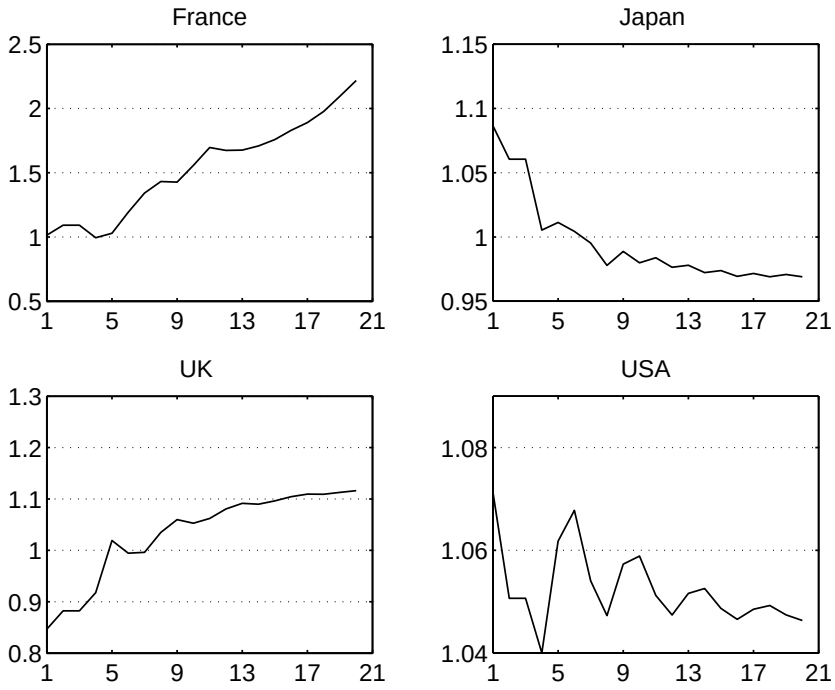
$x_t = [\Delta y_t \quad \Delta p_t \quad \Delta q_t]'$,

where y_t is the relative log GDP (home-foreign), p_t the difference of log price levels (home-foreign), q_t the log real exchange rate and $\Delta(1-L)$ the first difference operator.
The time series for France, Germany, Japan, the UK and the US from 1973:1 to 1998:4 are used to estimate 4 VARs relative to Germany. The data is described in section (3.5.3) of the appendix.

3.3.1 Specification

Section (3.5.4.3.1) shows the results of the Belsley, Kuh, and Welsch (1980) test of collinearity of relative logarithm and differences of relative logarithm series. This test detects near linear relations for (at least two) variables if the conditions index $\kappa(x)$ exceeds 30 and variance-decomposition proportions are greater than 50 for at least two variables. Only for the relative log specification of France ($\kappa(x) = 43, y - y^* = 1.00, q = 0.53$) there seems to be a near linear relationship between the relative output and the real exchange rate. For the difference specification there are no linear relationships detectable.
The specification of the model requires unit root tests for the involved time series. Therefore augmented Dickey-Fuller tests are performed. The lag length for these tests is determined by the Breusch-Godfrey LM test in such a way that no autocorrelation up to order 6 is detectable at a significance level of 5%. It follows theoretically from processes of the exogenous variables in (3.7) that the relative output and the real exchange rate follow unit root processes. The results of these two tests are listed in section (3.5.4.3.2). Only France's real exchange rate is at 1% level significantly stationary. The UK's relative price stationarity at 5% is rather strange as the graph in section (3.5.4.1) reveals. All other time series have unit roots. The lower part of the table shows that the first differences of all series are I(0) as the graphs in section (3.5.4.2) suggest.
Therefore specification in first differences is appropriate, if there is no cointegration. From the triangular structure of the long run solutions in equations

Figure 3.1: Real and nominal exchange rate response to real shocks



(3.8)-(3.10) $\tilde{y}_t = y(z_t), \tilde{q}_t = q(z_t, \delta_t), \tilde{p}_t = p(z_t, \delta_t, v_t)$ it follows that the variables should not be cointegrated, as every variable is driven by another independent process, whereas cointegration is equivalent to common trends in the relevant variables. Section (3.5.4.3.3) confirms it empirically, that there is no cointegration in the VAR systems and an error correction model is not necessary.
The lag length in the first difference VARs is determined by a sequential likelihood ratio test on the lag length starting from 12 lags. A lag length of 6 quarters for France, 4 for Japan, 5 for the UK and 4 for the USA seem to be appropriate.

3.3.2 Estimation, Identification and Interpretation

The identification procedure is given in section (3.5.2) of the appendix. The identified impulse responses are graphed in section (3.5.4.4). The forecast variance decomposition can be found in section (3.5.4.5). The ratios of nominal and real exchange rate impulse responses are calculated in section (3.5.4.6). The upper two graphs of the figures for each country reproduce the impulse response functions of the real and nominal exchange rates and the relative price to supply and demand shocks. The lower part graphs the ratio of the real and nominal exchange rate impulse responses. Values between 0 and 1 support Hypothesis (1). Values above 1 indicate a move of the nominal exchange rate in the right

Table 3.1: Long run impulse responses $F(1)$

	France			UK		
$y - y^*$	0.84	0.00	0.00	1.49	0.00	0.00
$p - p^*$	-2.47	4.56	2.14	-0.47	0.84	2.16
q	**1.22**	**-3.43**	0.00	**1.61**	**-4.96**	0.00
$s = q + p - p^*$	**-1.25**	**1.13**	2.14	**1.14**	**-4.12**	2.16
|q-s|	2.47	5.56		✓0.47	✓0.84	
	Japan			USA		
$y - y^*$	1.20	0.00	0.00	1.54	0.00	0.00
$p - p^*$	-0.09	-0.03	1.16	-0.43	0.78	1.46
q	**0.53**	**-4.32**	0.00	**-2.06**	**-5.55**	0.00
$s = q + p - p^*$	**0.44**	**-4.35**	1.16	**-2.49**	**-4.77**	1.46
|q-s|	✓0.09	✓0.03		(✓)0.43	✓0.78	
shocks	supply	demand	monetary	supply	demand	monetary

direction but too far so that the relative price level has to readjust. Values below zero indicate a move of the nominal exchange rate in the opposite direction of the real exchange rate effects.
The real and nominal exchange rate reaction ratios of Figure 3.1 aggregates over both real shocks. Table 3.1 shows the long run impulse responses. The snags mark impulse response functions which are in line with the hypothesis. The end of a nominal exchange rate adjustment is a small difference between the impulse response function of the nominal and real exchange rate.
The ranking according to this aim is Japan, USA, UK and France. Between Germany and Japan and Germany and USA a floating regime was in place over the analyzed period. The regime between German and UK was mostly a floating system whereas between Germany and France we had mostly a fixed exchange rate regime. These different regimes may explain the differences in the exchange rate adjustments.

3.4 Conclusion

The shock absorbing property of the nominal exchange rate can be clearly detected for Japan and the UK. The dollar exchange rate seems to overreact to supply shocks but is in line with the hypothesis for demand shocks. The nominal exchange rate vis à vis France does not behave as a shock absorber - quite the opposite.
A structural approach with estimated parameters and shocks of an explicit model is desirable for a more conclusive analysis.

3.5 Appendix

3.5.1 Flexible exchange rate solutions

3.5.1.1 Long run solutions

3.5.1.1.1 Real exchange rate The solution for the real exchange rate is obtained by plugging UIP (3.5) and the definition of the real exchange rate (3.6) into the IS curve (3.1).

$$\begin{aligned} y_t &= \eta(s_t - p_t) - \sigma[E_t(s_{t+1} - s_t) - E_t(p_{t+1} - p_t)] + d_t \\ &= \eta(s_t - p_t) - \sigma[E_t(q_{t+1}) - q_t] \qquad\qquad + d_t \end{aligned}$$

It follows that

$$\tilde{q}_t = \frac{y_t - d_t}{\eta + \sigma} + \frac{\sigma}{\eta + \sigma} E_t(\tilde{q}_{t+1}). \tag{3.20}$$

The method of undetermined coefficients is used to solve this stochastic difference equation. The guess solution is that q_t follows a linear process:

$$\tilde{q}_t = a_1 y_t + a_2 m_t + a_3 d_t + a_4 \delta_t. \tag{3.21}$$

Moving this equation one period ahead, inserting the exogenous processes (3.7) and taking expectations conditioned on information at time t, gives

$$E_t(\tilde{q}_{t+1}) = a_1 y_t + a_2 m_t + a_3(d_t - \gamma\delta_t).$$

Substitute this equation together with (3.21) into (3.20) to get

$$a_1 y_t + a_2 m_t + a_3 d_t + a_4 \delta_t = \frac{y_t - d_t}{\eta + \sigma} + \frac{\sigma}{\eta + \sigma}[a_1 y_t + a_2 m_t + a_3(d_t + \gamma\delta_t)],$$

which determines the coefficients:

$$a_1 = \frac{1}{\eta} = -a_3, \quad a_2 = 0, \quad a_4 = \frac{\gamma}{\eta}\frac{\sigma}{\eta + \sigma},$$

and the solution for the real exchange rate:

$$\tilde{q}_t = \frac{y_t - d_t}{\eta} + \frac{\gamma}{\eta}\frac{\sigma}{\eta + \sigma}\delta_t. \tag{3.22}$$

3.5.1.1.2 Relative price The solution for the relative price is obtained by substituting the UIP (3.5) into the LM curve (3.2), adding $\lambda\tilde{p}_t$ on both sides and adding and subtracting $\lambda E_t(\tilde{p}_{t+1})$ on the RHS.

$$\begin{aligned} \tilde{p}_t + \lambda\tilde{p}_t &= m_t - y_t + \lambda E_t(s_{t+1} - s_t) + \lambda\tilde{p}_t + \lambda E_t(\tilde{p}_{t+1}) - \lambda E_t(\tilde{p}_{t+1}) \\ (1 + \lambda)\tilde{p}_t &= m_t - y_t + \lambda E_t(\tilde{q}_{t+1} - \tilde{q}_t) + \lambda E_t(\tilde{p}_{t+1}) \end{aligned} \tag{3.23}$$

Then equation (3.22) is forwarded one period ahead, the exogenous processes are replaced by (3.7), equation (3.22) is subtracted, and conditional expectations are taken to arrive at

$$\begin{aligned}\tilde{q}_{t+1} - \tilde{q}_t &= \frac{z_{t+1} - \delta_{t+1} + \gamma\delta_t}{\eta} + \frac{\gamma}{\eta}\frac{\sigma}{\eta+\sigma}(\delta_{t+1} - \delta_t) \\ E_t(\tilde{q}_{t+1} - \tilde{q}_t) &= \frac{\gamma\delta_t}{\eta} - \frac{\gamma}{\eta}\frac{\sigma}{\eta+\sigma}\delta_t = \frac{\gamma}{\eta+\sigma}\delta_t,\end{aligned}$$

which is plugged into (3.23), giving a stochastic difference equation for $\tilde{p}_t$:

$$\tilde{p}_t = \frac{m_t - y_t}{1+\lambda} + \frac{\lambda\gamma}{(\eta+\sigma)(1+\lambda)}\delta_t + \frac{\lambda}{1+\lambda}E_t(\tilde{p}_{t+1}). \tag{3.24}$$

The guess solution for the method of undetermined coefficients is:

$$\tilde{p}_t = b_1 m_t + b_2 y_t + b_3 d_t + b_4 \delta_t. \tag{3.25}$$

Forwarding this equation one period ahead, using the definition of exogenous processes in (3.7) and taking expectations conditioned on information at time t gives

$$E_t(\tilde{p}_t) = b_1 m_t + b_2 y_t + b_3(d_t - \gamma\delta_t).$$

Substitute this equation and (3.25) into (3.24) to get

$$\begin{aligned}b_1 m_t + b_2 y_t + b_3 d_t + b_4\delta_t &= \frac{m_t - y_t}{1+\lambda} + \frac{\lambda\gamma}{(\eta+\sigma)(1+\lambda)}\delta_t \\ &+ \frac{\lambda}{1+\lambda}b_1 m_t + b_2 y_t + b_3(d_t - \gamma\delta_t).\end{aligned}$$

Equate the coefficients on both sides to get

$$b_1 = 1 = -b_2, \quad b_3 = 0, \quad b_4 = \frac{\lambda\gamma}{(1+\lambda)(\eta+\sigma)}.$$

The solution for the price difference is

$$\tilde{p}_t = m_t - y_t + \alpha\gamma\delta_t \tag{3.26}$$

$$\alpha = \frac{\lambda}{(1+\lambda)(\eta+\sigma)}. \tag{3.27}$$

3.5.1.2 Short run solutions

3.5.1.2.1 Relative price The solution for the relative price is obtained, if the long run solution for the relative price (3.26) and the exogenous processes (3.7) are substituted into the price adjustment equation (3.4) and $v_t - z_t + \alpha\gamma\delta_t$ is added and subtracted.

$$\begin{aligned}p_t &= \theta(m_{t-1} + v_t - y_{t-1} - z_t + \alpha\gamma\delta_t) + (1-\theta)E_{t-1}(m_{t-1} + v_t - y_{t-1} - z_t + \alpha\gamma\delta_t) \\ &= m_{t-1} - y_{t-1} + \theta(v_t - z_t + \alpha\gamma\delta_t) \\ &= \tilde{p}_t - (1-\theta)(v_t - z_t + \alpha\gamma\delta_t)\end{aligned} \tag{3.28}$$

3.5.1.2.2 Real exchange rate To obtain the solution for the real exchange rate the short run solution for the price level (3.28), the exogenous processes (3.7), the IS curve (3.1), and UIP (3.5) is substituted into the LM curve (3.2) and $\lambda E_t(p_{t+1} - p_t)$ is added and subtracted.

$$m_t - \tilde{p}_t + (1-\theta)(v_t - z_t + \alpha\gamma\delta_t) = d_t + \eta q_t - (\sigma - \lambda)E_t(q_{t+1} - q_t) - \lambda E_t(Pt + 1 - p_t) \tag{3.29}$$

From equation (3.28) and (3.26) we get

$$E_t(p_{t+1} - p_t) = -\alpha\gamma\delta_t + (1-\theta)(v_t - z_t + \alpha\gamma\delta_t).$$

Substituting this equation, and the short run solution for the relative price (3.28) into (3.29) we arrive at a stochastic difference equation for the real exchange rate.

$$(\eta + \sigma + \lambda)q_t = y_t - d_t + (1-\theta)(v_t - z_t) - \theta(1+\lambda)\alpha\gamma\delta_t + (\sigma + \lambda)E_t(q_{t+1}). \tag{3.30}$$

The guess solution for the method of undetermined coefficients is

$$q_t = c_1 y_t + c_2 d_t + c_3 \delta_t + c_4 v_t + c_5 z_t.$$

Forwarding this equation one period ahead, substituting the exogenous processes (3.7) and taking the conditional expectations we arrive at

$$E_t(q_{+1}) = c_1 y_t + c_2(d_t - \gamma z_t).$$

Substituting the last two equations into (3.30)

$$(\eta + \sigma + \lambda)c_1 y_t + c_2 d_t + c_3 \delta_t + c_4 v_t + c_5 z_t =$$

$$y_t - d_t + (1-\theta)(v_t - z_t) - \theta(1+\lambda)\alpha\gamma\delta_t + (\sigma + \lambda)(c_1 y_t + c_2(d_t - \gamma z_t))$$

and equating the coefficients of the variables, we get the determined coefficients:

$$c_1 = \frac{1}{\eta} = -c_2, \quad c_3 = \frac{\gamma(\sigma + \lambda) - \eta\alpha\gamma\theta(1+\lambda)}{\eta(\eta + \sigma + \lambda)}, \quad c_4 = \frac{(1-\theta)(1+\lambda)}{\eta + \sigma + \lambda} = -c_5,$$

and the solution for the real exchange rate

$$q_t = \frac{y_t - d_t}{\eta} + \frac{\gamma(\sigma + \lambda) - \eta\alpha\gamma\theta(1+\lambda)}{\eta(\eta + \sigma + \lambda)} d_t + \frac{(1-\theta)(1+\lambda)}{\eta + \sigma + \lambda}(v_t - z_t).$$

Equations (3.27) and (3.22) can be used to express this solution as a function of news

$$q_t = \tilde{q}_t + \frac{(1-\theta)(1+\lambda)}{\eta + \sigma + \lambda}(v_t - z_t + \alpha\gamma\delta_t) \tag{3.31}$$

3.5.1.2.3 Output The solution for the relative output is obtained from the aggregate demand

$$y_t^d = \eta q_t - \sigma E_t(\Delta q_{t+1}) + d_t.$$

Take the conditional expectation of (3.31)

$$E_t(\Delta q_{t+1}) = \frac{\gamma}{\eta+\sigma}\delta_t - \frac{(1-\theta)(1+\lambda)}{\eta+\sigma+\lambda}(v_t - z_t + \alpha\gamma\delta_t)$$

and substitute this equation together with the short run solution for the relative price (3.31) into the aggregate demand, it follows that

$$y_t^d = y_t^s + \frac{(1-\theta)(1+\lambda)(\eta+\sigma)}{\eta+\sigma+\lambda}(v_t - z_t + \alpha\gamma\delta_t). \tag{3.32}$$

3.5.2 Identification

Identification of the structural VAR follows Blanchard and Quah (1989) by using the theory based long run restriction derived in section (3.2.1). The estimated VAR includes three variables:

$$x_t = [\Delta y_t \quad \Delta p_t \quad \Delta q_t]',$$

where y_t is the difference of log GDP (home-foreign), q_t the real exchange rate, p_t the difference of log price levels (home-foreign) and Δ the first difference operator $(1-L)$.
The first step is to estimate the reduced form VAR with p lags:

$$\begin{aligned} x_t &= A_1 x_{t-1} + A_2 x_{t-2} + \ldots + A_p x_{t-p} + e_t \\ &= (I - \sum_{j=1}^{p} A_j L^j)^{-1} e_t = \sum_{j=0}^{\infty} C_j L^j e_t = C(L) e_t \end{aligned}$$

In order to obtain impulse response functions (coefficients of the structural MA representation) the estimated reduces VAR has to be inverted into its MA representation. This is done by exploiting the identity

$$I = (\sum_{j=1}^{\infty} C_j L^j)(I - \sum_{j=0}^{p} A_j L^j) = \sum_{j=0}^{\infty}(C_j - \sum_{k=1}^{j} C_{j-k} A_k) L^j$$

and equating the coefficients of the same powers of the lag operator.

$$\begin{aligned}
C_0 &= I_0 \\
C_1 &= A_1 \\
C_2 &= C_1A_1 + A_2 \\
C_3 &= C_2A_1 + C_1A_2 + A_3 \\
C_4 &= C_3A_1 + C_2A_2 + C_1A_3 + A_4 \\
&\vdots \\
C_k &= \sum_{j=1}^{k} C_{k-j}A_j.
\end{aligned}$$

The coefficients C_k establish the MA representation of the reduced model.

$$x_t = \sum_{j=0}^{\infty} C_j L^j e_t = C(L)e_t \tag{3.33}$$

where $E(e_te_t') = \Sigma$, $C_0 = I$, $C(L) = \sum_{j=0}^{\infty} C_jL^j$.
The variables are driven by structural shocks $\epsilon_t = (z_t, \delta_t, v_t)'$, though the structural model and its MA representation is

$$x_t = \sum_{j=0}^{\infty} F_j L^j \epsilon_t = F(L)\epsilon_t. \tag{3.34}$$

The reduced (3.33) and structural (3.34) form reveal the connection between both disturbances:

$$e_t = F_0\epsilon_t, \quad e_{t-j} = F_0\epsilon_{t-j}, \quad C_je_{t-j} = C_jF_0\epsilon_{t-j} = F_j\epsilon_{t-j}, \quad F_j = C_jF_0$$

$$F(1) = C(1)F_0, \quad C(1)\Sigma C(1)' = F(1)F(1)', \quad F(1) = \sum_{j=0}^{\infty} F_j, \quad C(1) = \sum_{j=0}^{\infty} C_j$$

To identify the structural shocks matrix F_0 has to be determined which together with the coefficient matrices C_j gives the impulse responses F_j.
First I want to show a matrix version of the identification process and then a procedure relying on the elements of F_0. The purpose is to show that matrix derivation obscures implicit assumptions about the signs of matrix $F0$ which are usually not discussed as statements like 'the signs of the estimated impulse response functions match theoretical considerations' witness.
The matrix version works as follows. First, we need a Cholesky decomposition of the estimated MA coefficients of the reduced form. The covariance matrix is then transformed into $C(1)\Sigma C(1) = HH'$ where H is lower triangular. Since $e_te_t' = F_0\epsilon_t\epsilon_f'F_0' \leftrightarrow \Sigma = F_0F_0' \leftrightarrow C(1)\Sigma C(1)' = C(1)F_0F_0C(1)' = F(1)F(1)'$ it follows

that $F(1) = H$ and therefore $F_0 = C(1)^{-1}F(1) = C(1)H$. The RHS $C(1)H$ is completely determined by estimation results and gives therefore the identification matrix F_0. Equating $A(1)$ to H uses the first assumption of $E(\epsilon_t\epsilon_t') = I$, i.e. unit variance orthogonal structural shocks. Since H is a lower triangular matrix, the long run restrictions are implemented as a by product. The crucial point is that it seems as if there is a unique solution for F_0 but in fact there are 8, distinguished by the signs of its elements.
To see this point more clearly a derivation relying on the elements of F_0 is presented. Matrix F_0 has 9 elements which have to be identified. The first 6 elements are derived from the assumption that structural shocks have unit variance and are uncorrelated. Therefore we can obtain from $e_te_t' = F_0\epsilon_t\epsilon_f'F_0'$ the equation $\Sigma = F_0F_0'$. This gives us only 6 equations for the 9 unknown elements of matrix F_0, since the variance matrix Σ is symmetric. g_{ij} and f_{ij} denote the elements of matrix G and F_0, respectively.

$$\left[\begin{array}{c|c|c} g_{11}^2 & g_{11}g_{12} & g_{11}g_{13} \\ g_{11}g_{12} & g_{12}^2+g_{22}^2 & g_{12}g_{13}+g_{22}g_{23} \\ g_{11}g_{13} & g_{12}g_{13}+g_{22}g_{23} & g_{13}^2+g_{23}^2+g_{33}^2 \end{array}\right] \stackrel{!}{=} \tag{3.35}$$

$$\left[\begin{array}{c|c|c} f_{11}^2+f_{12}^2+f_{13}^2 & f_{11}f_{21}+f_{12}f_{22}+f_{13}f_{23} & f_{11}f_{31}+f_{12}f_{32}+f_{13}f_{33} \\ f_{11}f_{21}+f_{12}f_{22}+f_{13}f_{23} & f_{21}^2+f_{22}^2+f_{23}^2 & f_{21}f_{31}+f_{22}f_{32}+f_{23}f_{33} \\ f_{11}f_{31}+f_{12}f_{32}+f_{13}f_{33} & f_{21}f_{31}+f_{22}f_{32}+f_{23}f_{33} & f_{31}^2+f_{32}^2+f_{33}^2 \end{array}\right]$$

Another 3 restrictions are derived from the long run implications of the model. According to section (3.2.1) neither demand (IS) nor monetary (LM) shocks affect the output *level* in the long run, furthermore the real exchange rate *level* is unaffected by monetary (LM) shocks in the long run. Since these assumptions are concerned with effects of shocks on the level of the variables and the system is estimated in differences they have to be formulated in relation to the infinite cumulated sum of the impulse response functions, i.e. the long run impulse responses. $c_{ij}(1)$ and $f_{ij}(1)$ denote the elements of matrix $C(1) = \sum_{j=0}^{\infty} C_j$ and $F(1) = \sum_{j=0}^{\infty} F_j$, respectively. From $F(1) = C(1)F_0$ the last restrictions are derived.

$$\begin{aligned} f_{12}(1) &\stackrel{!}{=} 0 = c_{11}(1)f_{12} + c_{12}(1)f_{22} + c_{13}(1)f_{32} \\ f_{13}(1) &\stackrel{!}{=} 0 = c_{11}(1)f_{13} + c_{12}(1)f_{23} + c_{13}(1)f_{33} \\ f_{33}(1) &\stackrel{!}{=} 0 = c_{31}(1)f_{13} + c_{32}(1)f_{23} + c_{33}(1)f_{33}. \end{aligned} \tag{3.36}$$

These 3 equations give the last 3 restrictions on f_{ij} to identify F_0 exactly and to calculate the impulse responses $F_j = C_jF_0$.
To see the 8 solutions for F_0 note that equation (3.35) is quadratic in the elements of F_0. The sign of f_{11} can be changed without affecting the solution if according to the equations (3.36) the signs of f_{21} and f_{31}) are changed too. This won't

change equation (3.35) since f_{11}, f_{21}, f_{31} occur only as products among themselves. Accordingly solving equations (3.35) and (3.36) gives $2^3 = 8$ solutions. As an example the solutions for France's F_0 are given in the following table. They are obtained by different starting values for the numerical solution.

$\Delta(y - y^*)$	-0.65	0.44	0.35	0.65	0.44	0.35
$\Delta(p - p^*)$	0.36	0.38	0.12	-0.36	0.38	0.12
Δq	-0.03	-0.96	1.11	0.03	-0.96	1.11
	supply	demand	money			
	-0.65	0.44	-0.35	0.65	0.44	-0.35
	0.36	0.38	-0.12	-0.36	0.38	-0.12
	-0.03	-0.96	-1.11	0.03	-0.96	-1.11
	-0.65	-0.44	0.35	0.65	-0.44	0.35
	0.36	-0.38	0.12	-0.36	-0.38	0.12
	-0.03	0.96	1.11	0.03	0.96	1.11
	-0.65	-0.44	-0.35	0.65	-0.44	-0.35
	0.36	-0.38	-0.12	-0.36	-0.38	-0.12
	-0.03	0.96	-1.11	0.03	0.96	-1.11

In terms of identification this means that 3 additional assumptions on signs are needed. For each column of F_0 one sign has to be determined and the other two follow from the estimated F_0. For the above identification matrix this means that assuming a positive effect of a supply shock on the relative output growth implies a negative on relative inflation and a real depreciation.
From the impulse response functions the forecast error variance decompositions can be calculated, simplified by the identifying assumptions of the unit variance and the orthogonality of the structural shocks. The proportion of the variance of i-th variable accounted for by the j-th shock at horizon h ($R^2_{ij,h}$) is

$$R^2_{ij,h} = \frac{\sum_{k=0}^{h-1} f^2_{ij,k}}{\sum_{m=1}^{n} \sum_{k=0}^{h-1} f^2_{im,k}}$$

3.5.3 Data

GDP volume (seasonally adjusted), consumer price indices and nominal exchange rates are retrieved from the IMF *International Financial Statistics* (IFS) database. The sample covers the first quarter of 1973 until the fourth quarter of 1998 and includes 104 observations. IFS exchange rate figures are monthly average spot rates of national currencies per dollar (NC/$) which are transformed to quarterly averages of German marks to national currencies (DM/NC) since Germany is taken to be the home country. IFS values of Japan's GDP are apparently not seasonally adjusted from 1973:1 until 1979:1 therefore the standard U.S. Bureau of the Census adjustment method X-11 is applied. Germany's GDP

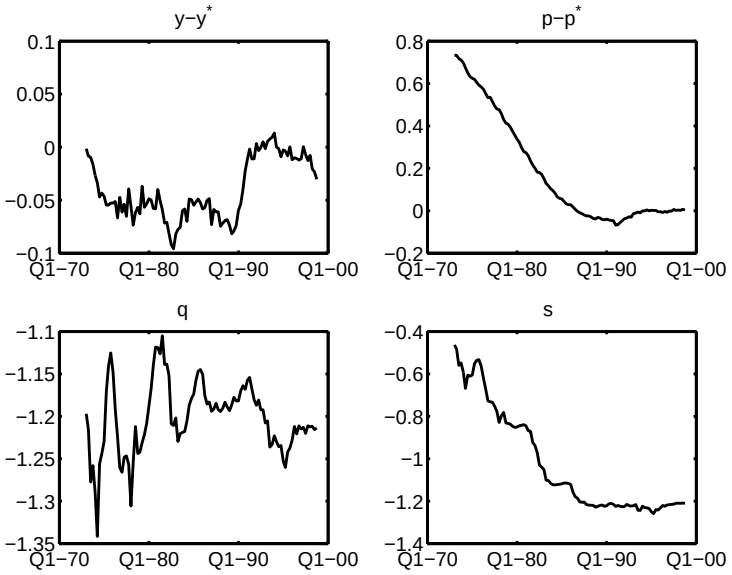
time series exhibits a level break in 1990:1 due to unification. The pre-unification index is transformed so that growth rates are preserved. The growth rate of 1989:4 to 1990:1 is assumed to be the mean of the last 4 growth rates. The IFS codes of the time series are:

	consumer prices	real GDP	nominal exchange rates
France	13264...ZF...	13299BVRZF...	132..AH.ZF...
Germany	13464...ZF...	13499BVRZF...	134..AH.ZF...
Japan	15864...ZF...	15899BVRZF...	158..AH.ZF...
United Kingdom	11264...ZF...	11299BVRZF...	112..AH.ZF...
United States	11164...ZF...	11199BVRZF...	111..AH.ZF...

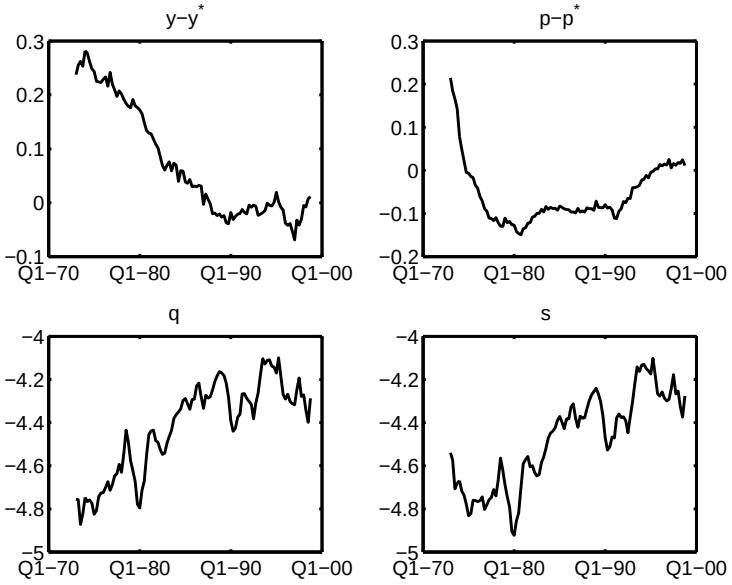
3.5.4 Table and Figures

3.5.4.1 Time Series: Relative Logarithms

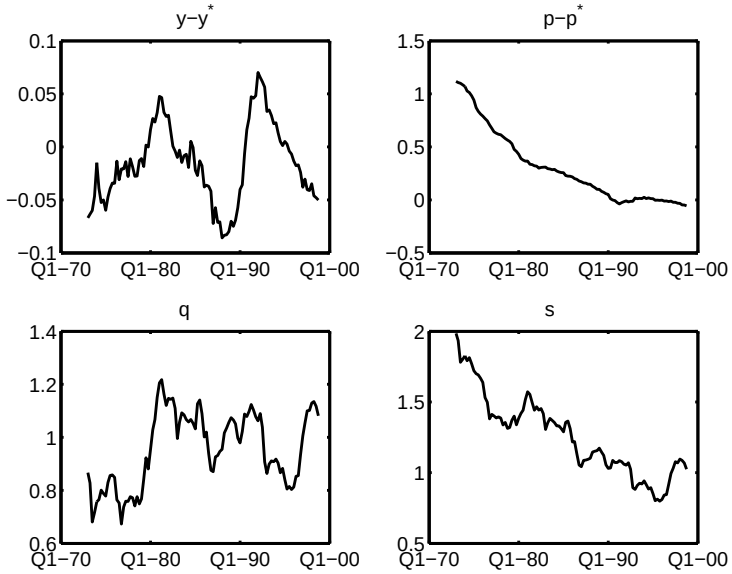
Germany - France



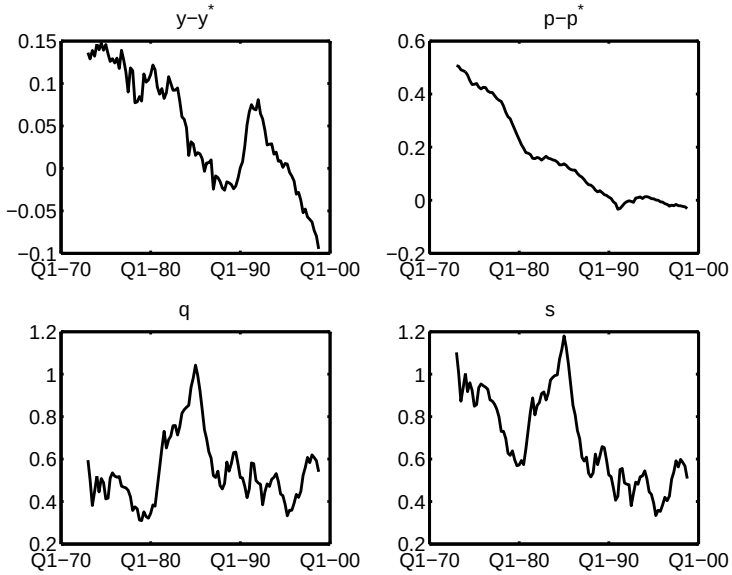
Germany - Japan



Germany - UK

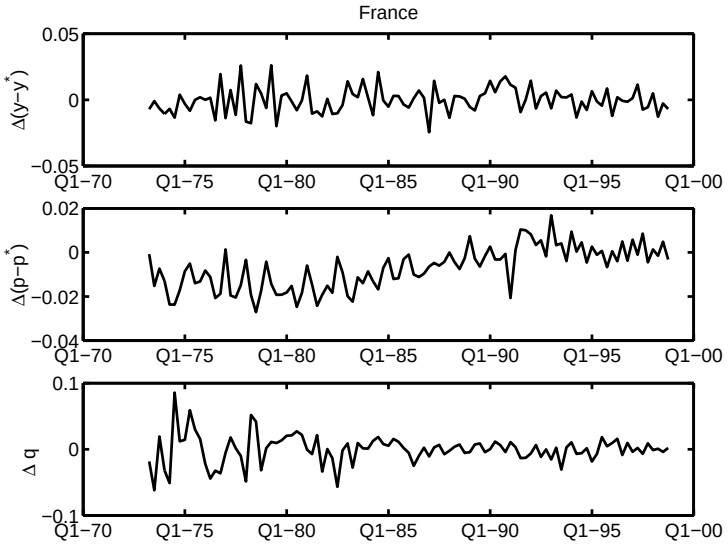


Germany - USA

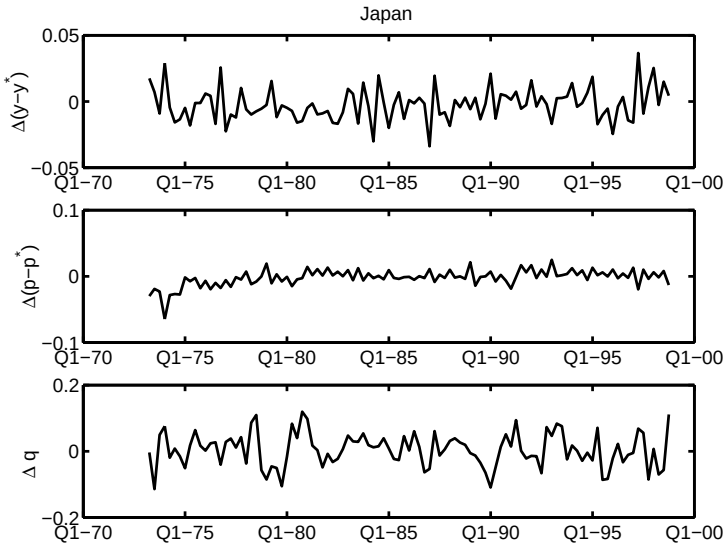


3.5.4.2 Time Series: First Difference of Relative Logarithms

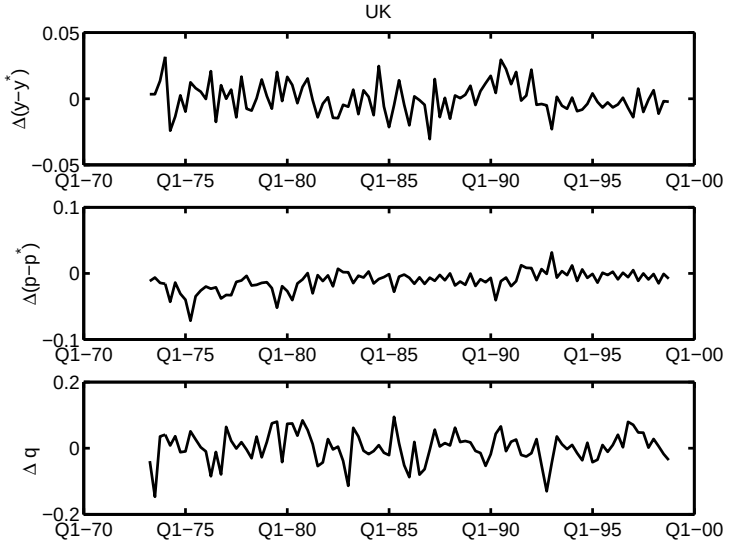
Germany - France



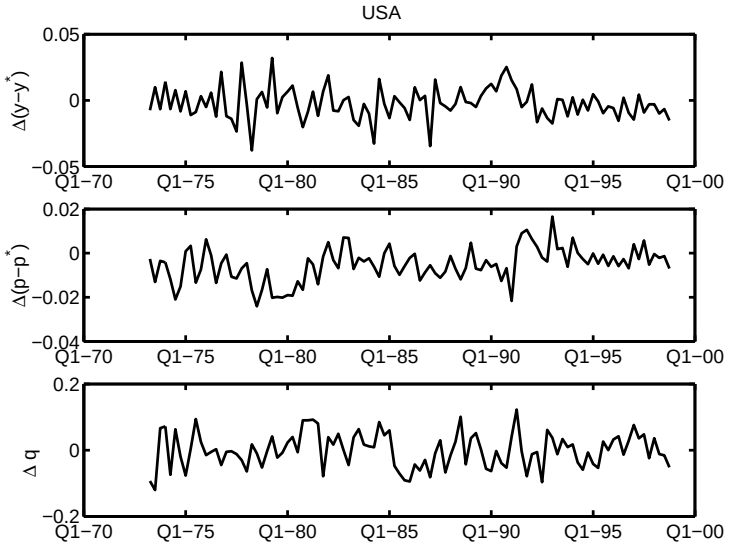
Germany - Japan



Germany - UK



Germany - USA



3.5.4.3 Specification Tests

3.5.4.3.1 Collinearity

$\kappa(x)$	$y-y^*$	$p-p^*$	q	$\kappa(x)$	$\Delta(y-y^*)$	$\Delta(p-p^*)$	Δq
France							
1	0.00	0.00	0.29	1	0.00	0.00	0.96
5	0.00	0.97	0.18	2	0.00	1.00	0.04
43	1.00	0.03	0.53	2	1.00	0.00	0.00
Japan							
1	0.00	0.00	0.40	1	0.00	0.00	0.95
44	0.92	0.02	0.17	4	0.45	0.32	0.01
65	0.08	0.98	0.43	5	0.55	0.68	0.03
UK							
1	0.00	0.01	0.53	1	0.00	0.00	0.88
3	0.00	0.89	0.46	3	0.01	0.94	0.10
28	1.00	0.10	0.01	4	0.99	0.06	0.02
USA							
1	0.00	0.00	0.54	1	0.00	0.00	0.99
3	0.00	0.24	0.46	4	1.00	0.00	0.01
17	1.00	0.76	0.00	6	0.00	1.00	0.00

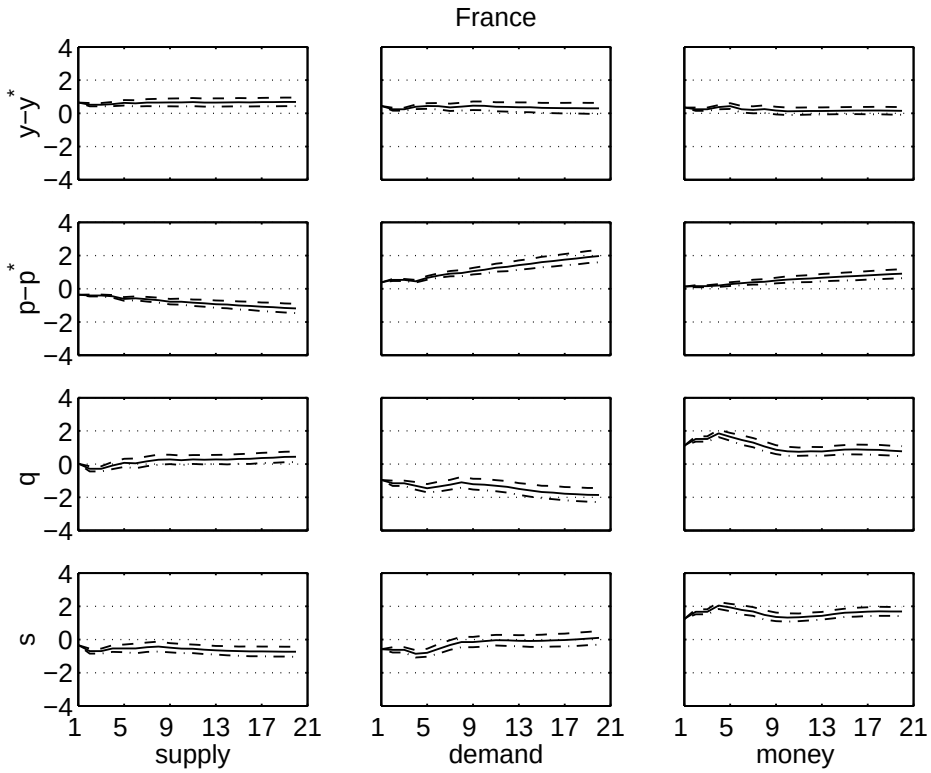
3.5.4.3.2 Augmented Dickey-Fuller

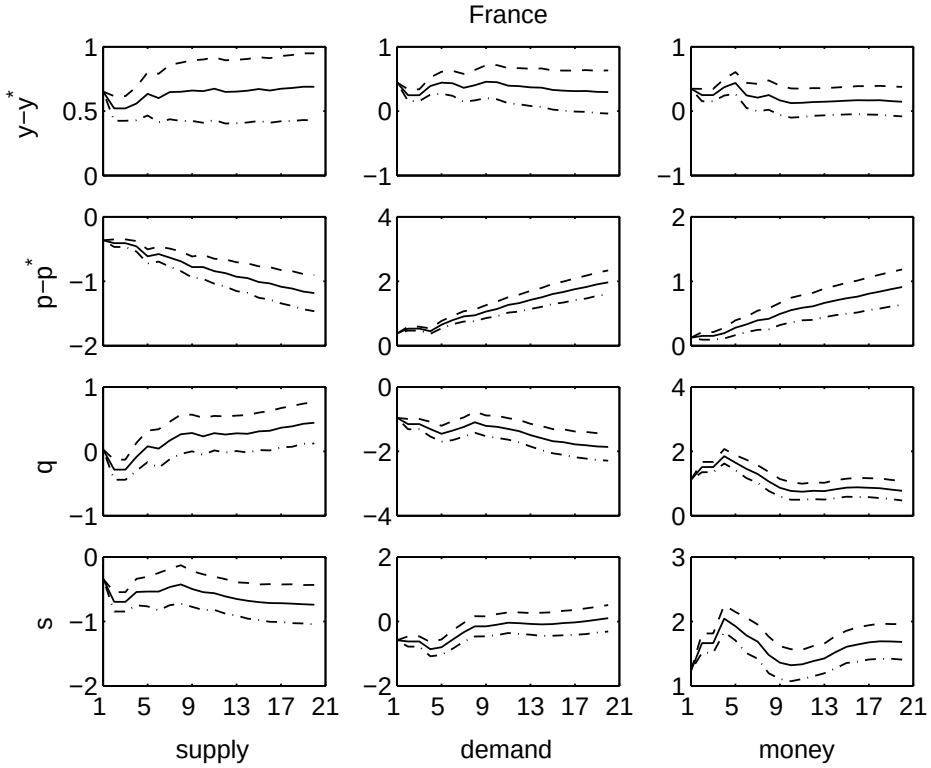
	Test statistic			Lag length		
	$y-y^*$	$p-p^*$	q	$y-y^*$	$p-p^*$	q
France	-1.925	-1.819	-5.062	4	6	3
Japan	-2.559	-1.177	-1.752	3	7	4
UK	-2.843	-3.193	-2.603	4	7	1
USA	-0.779	-1.786	-2.426	4	5	3
Signif.	1%	5%	10%			
Crit.Value	-3.439	-2.915	-2.584			
	$\Delta(y-y^*)$	$\Delta(p-p^*)$	Δq	$\Delta(y-y^*)$	$\Delta(p-p^*)$	Δq
France	-12,436	-3,548	-6,844	0	1	1
Japan	-7,860	-3,689	-6,613	1	1	1
UK	-6,349	-3,471	-3,321	1	1	1
USA	-4,940	-5,010	-4,255	2	1	2
Signif.	1%	5%	10%			
Crit.Value	-3.439	-2.915	-2.584			

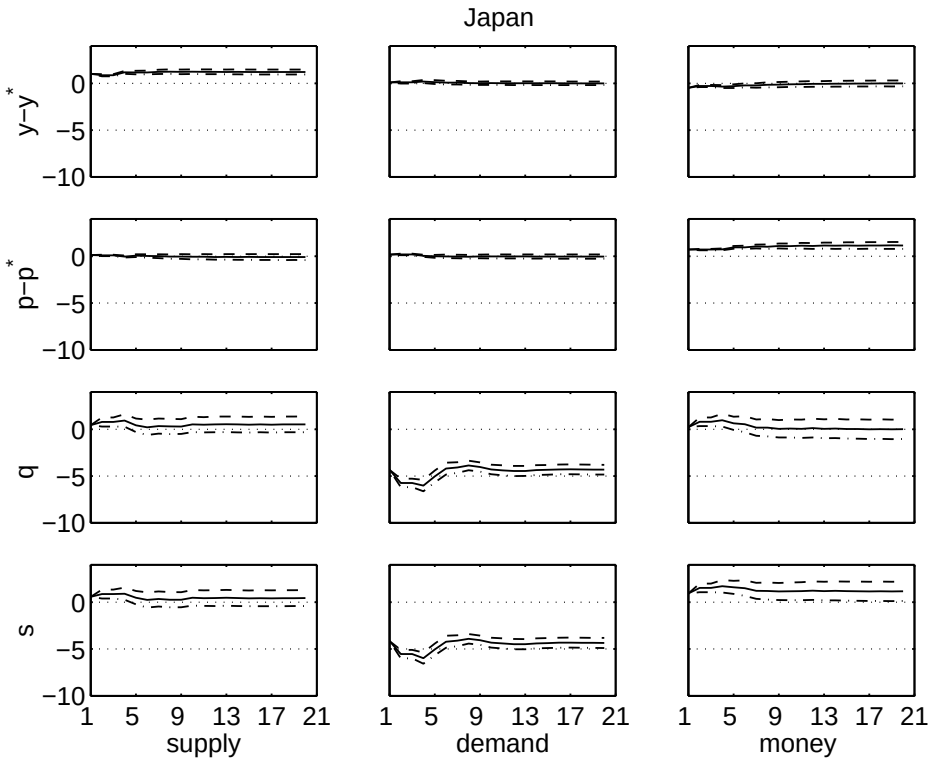
3.5.4.3.3 Cointegration Tests

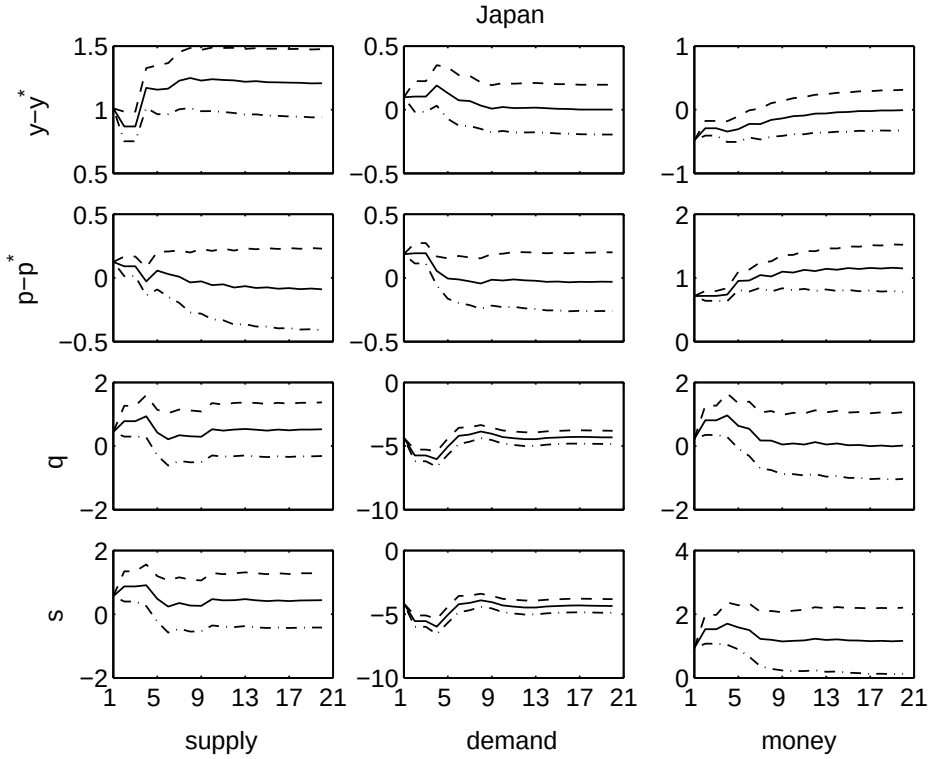
France			UK		
	Trace Statistic	Critical Value 95%		Trace Statistic	Critical Value 95%
r <= 0	27.539	35.012	r <= 0	24.52	35.012
r <= 1	9.184	18.398	r <= 1	9.925	18.398
r <= 2	1.901	3.841	r <= 2	4.692	3.841
	Eigenvalue Statistic	Critical Value 95%		Eigenvalue Statistic	Critical Value 95%
r <= 0	18.354	24.252	r <= 0	14.595	24.252
r <= 1	7.283	17.148	r <= 1	5.234	17.148
r <= 2	1.901	3.841	r <= 2	4.692	3.841
Japan			USA		
	Trace Statistic	Critical Value 95%		Trace Statistic	Critical Value 95%
r <= 0	26.513	35.012	r <= 0	25.55	35.012
r <= 1	6.947	18.398	r <= 1	11.965	18.398
r <= 2	0.019	3.841	r <= 2	3.194	3.841
	Eigenvalue Statistic	Critical Value 95%		Eigenvalue Statistic	Critical Value 95%
r <= 0	19.566	24.252	r <= 0	13.585	24.252
r <= 1	6.928	17.148	r <= 1	8.771	17.148
r <= 2	0.019	3.841	r <= 2	3.194	3.841

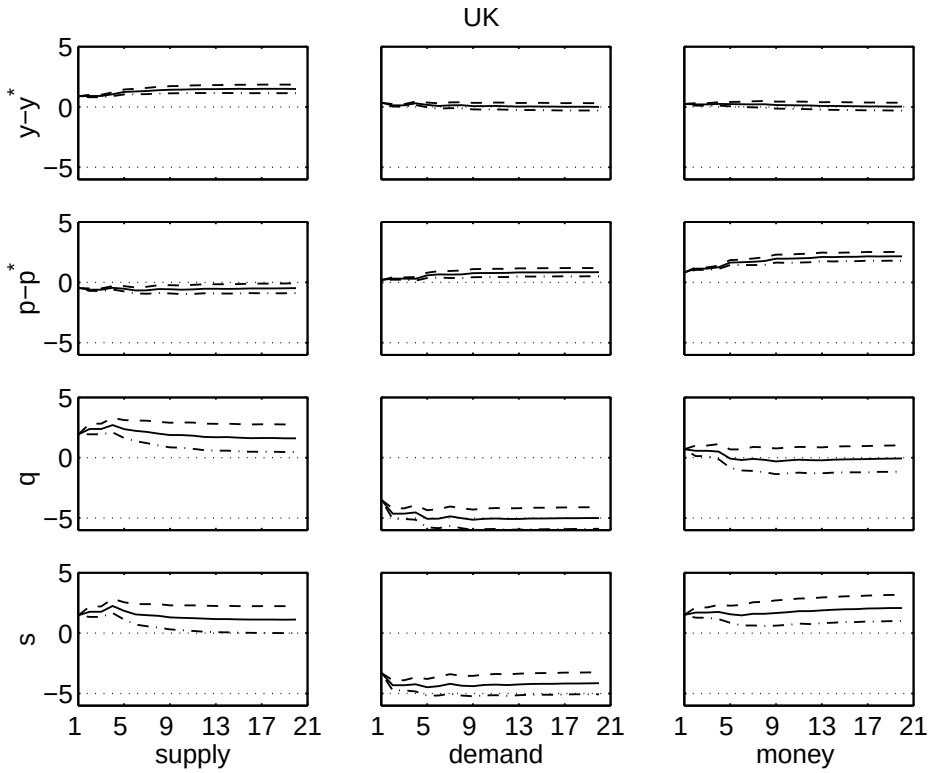
3.5.4.4 Impulse Response Functions

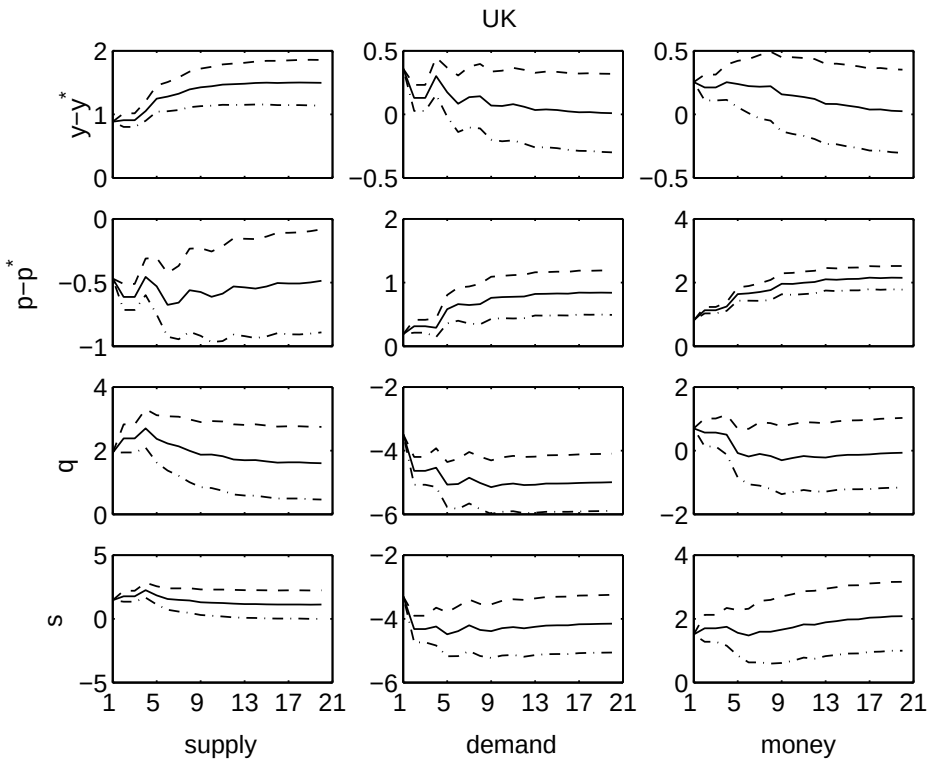


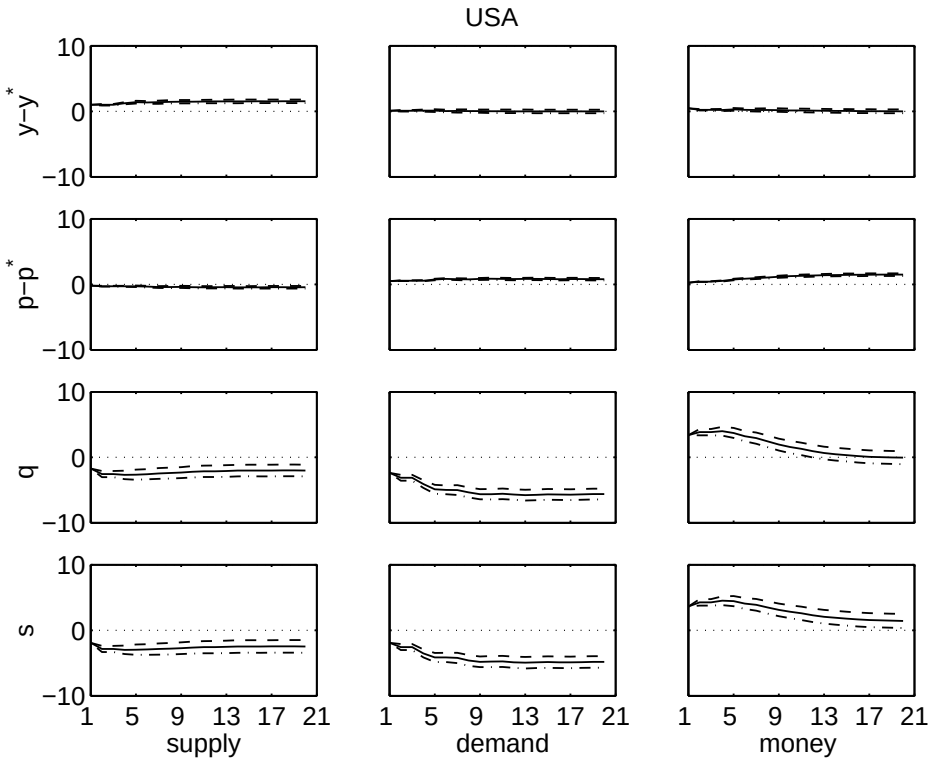


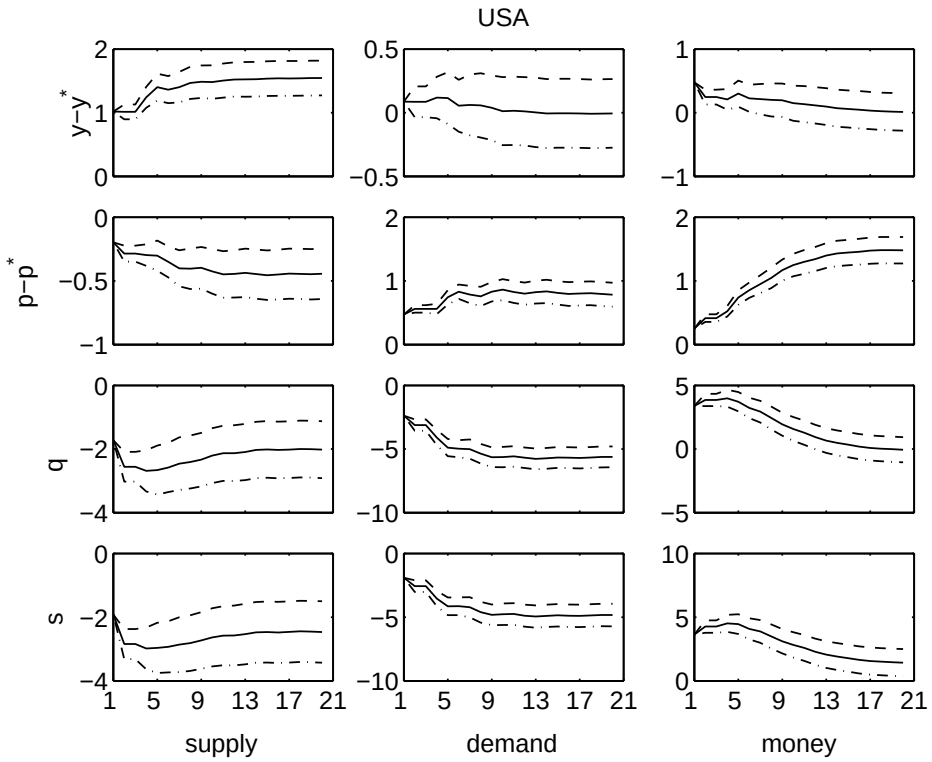

Japan
y–y*
p–p*
q
s
supply
demand
money


Japan
y–y*
p–p*
q
s
supply
demand
money

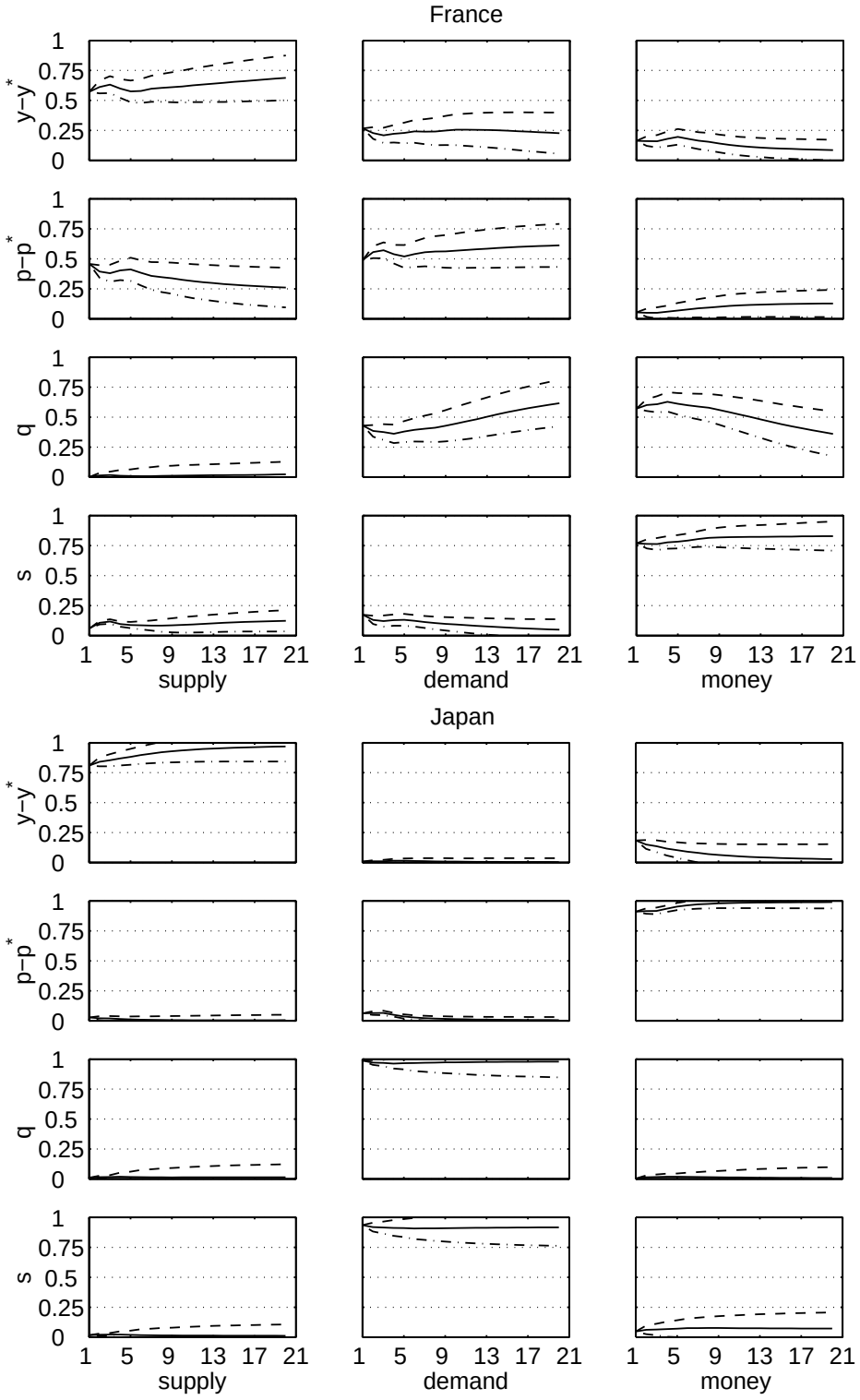

UK
y–y*
p–p*
q
s
supply
demand
money

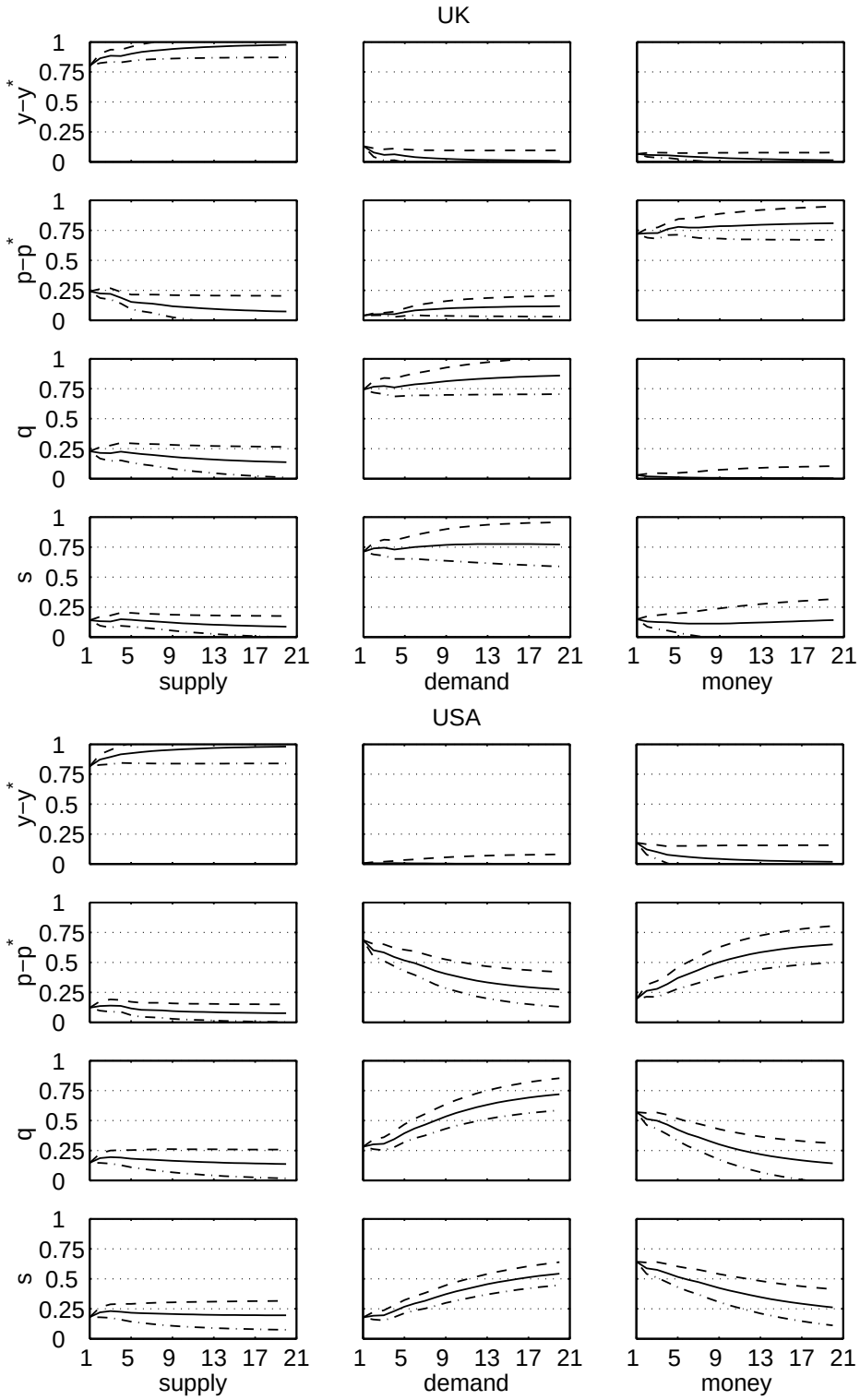

UK
y–y*
p–p*
q
s
supply
demand
money


USA
y–y*
p–p*
q
s
supply
demand
money

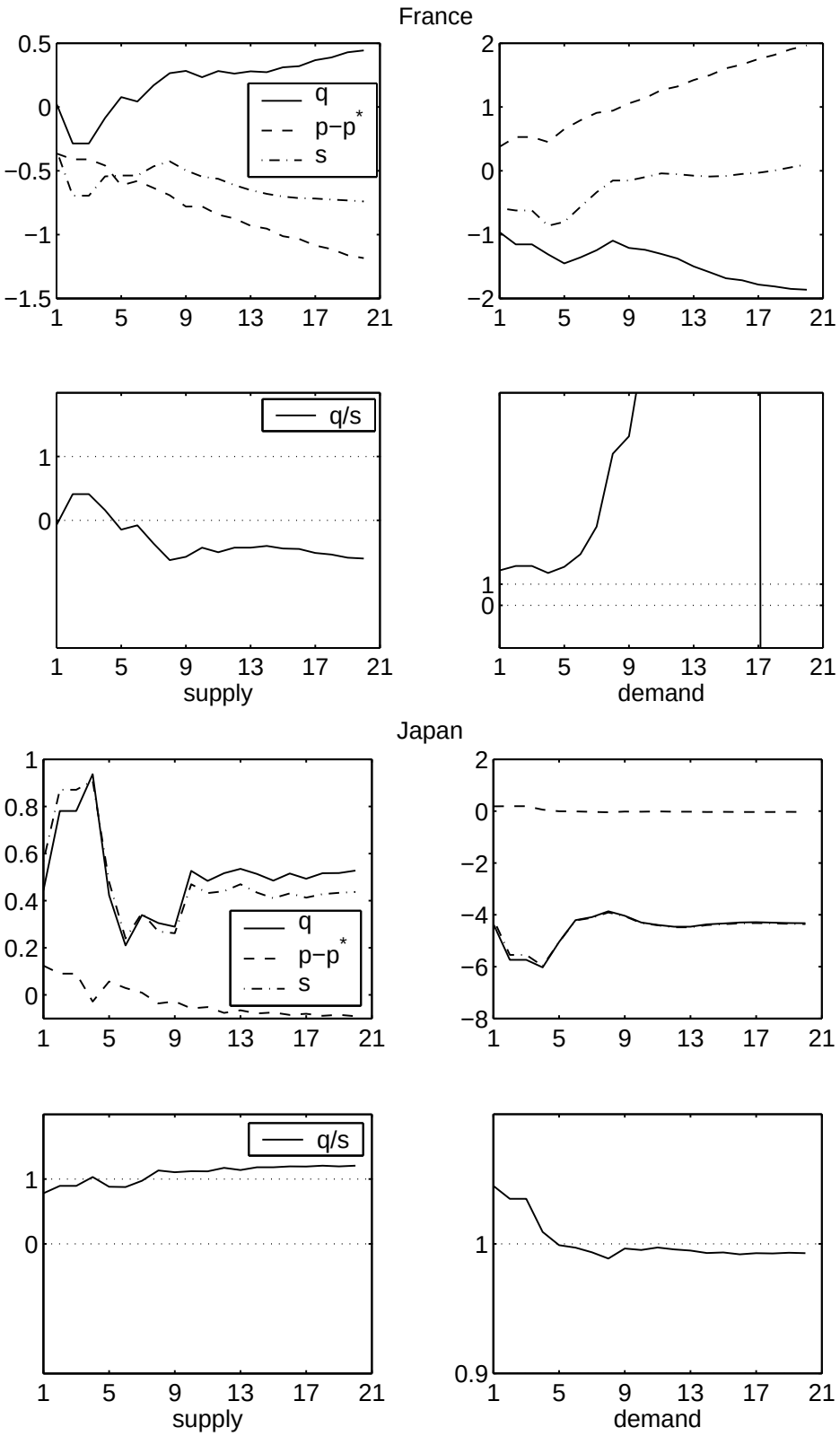

USA
y–y*
p–p*
q
s
supply
demand
money

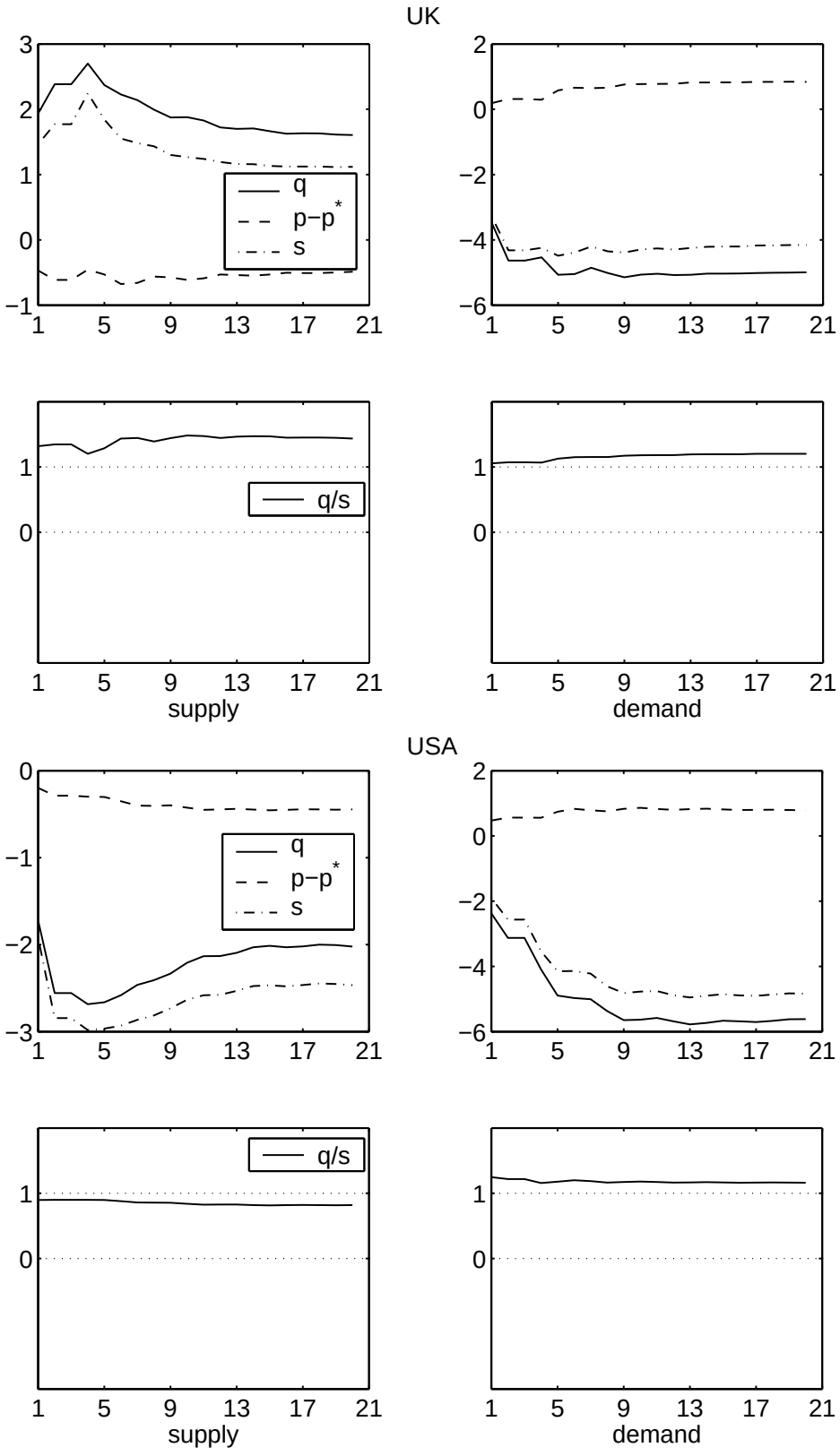
3.5.4.5 Forecast Error Variance Decomposition




UK
y*–y
p*–p
q
s
1
0.75
0.5
0.25
0
1 5 9 13 17 21
supply
demand
money
USA
y*–y
p*–p
q
s
1
0.75
0.5
0.25
0
1 5 9 13 17 21
supply
demand
money

3.5.4.6 Ratio of Real and Nominal Exchange Rate Impulse Response




UK
q
p−p*
s
q/s
supply
demand
USA
q
p−p*
s
q/s
supply
demand

Chapter 4

Solving, Estimating and Selecting Nonlinear Dynamic Economic Models without the Curse of Dimensionality

4.1 Introduction

Linear or linearized models are a drawback for policy evaluation and recommendation. They cannot account for important welfare channels under certainty equivalence when risk aversion and thus higher moments matter. Certainty equivalence describes the property of a policy to be independent of shock variances. Asset pricing models relate prices to risk measured by variances. The optimal currency area literature identifies the relative shock size as an important criterium for countries to join a currency union. The closely related theory of optimal exchange rate regimes discusses the optimal trade off between risks concerning prices, exchange rates or business cycles. It shapes the optimal policy as well as historical policy evaluations. The new open economics initiated by Obstfeld and Rogoff (1995) or the modern monetary economics, for example in Walsh (2001), take the nonlinearities into account when deriving optimal policies. Examples of spurious welfare effects are discussed in Kim and Kim (2003), Kollmann (2002) and Kollmann (2003). In positive economics nonlinear models help for example to account for business cycles features as in Morley and Piger (2005). My work provides a Bayesian estimation framework for general nonlinear macroeconomic models with rational expectations.

The estimation consists of three main steps. The first step is to solve the model and the second to evaluate the implied likelihood. The last step is a Metropolis-Hastings algorithm which generates a sample from the posterior density of structural parameters. The model can be linear or not, with rational expectations

or more general learning processes, Gaussian distributed shocks or not, observed and unobserved variables, fixed or time varying parameters. A model selection criterion selects parameter estimates whether models are nested or not, true or quasi-true. The generated sample from the posterior density can be used for a density estimate of a variable of interest. This can be any nonlinear function of observables and unobservables like a specification test statistic or any other economic variable. Most interesting are predictions and welfare measures for future or counterfactual policy simulations.

Beside the statistical difficulties to estimate a density, economic models are difficult to solve. The problem is to derive the optimal policy from a nonlinear functional equation with an expectations operator. Approximating the solution of a model is now well understood and documented in books like Judd (1998) or Marimon and Scott (1999). An applied solution approach with mixed discrete and continuous variables can be found in Miranda and Fackler (2002). Aruoba, Fernádez-Villaverde, and Rubio-Ramírez (2003) have recently compared the accuracy of available approaches for the same model as I use in the following.

Extensions of the nonlinear solution methods towards a likelihood based estimation are rare and the only one I know can be found in Fernádez-Villaverde and Rubio-Ramírez (2004b). They approximate the model solution by finite elements, rational expectations are evaluated by Kronecker Gaussian quadrature, the likelihood of the implied nonlinear state space model is calculated by a bootstrap filter and the Metropolis-Hastings algorithm with random walk innovations is used to draw random vectors from the posterior of structural parameters. This approach is very general and can handle any nonlinearity and shock distribution of the model. In particular nonsmooth policy functions can be approximated which may result from inequality constraints or min and max functions in the model formulation. The most urgent problem is that even with Fortran code they need around four days to estimate the smallest possible model and computationally more efficient methods are needed. The approach I present is related to their framework and extends and varies it in several dimensions with emphasis on faster algorithms for a general class of models.

The contribution of this work is to introduce the Smolyak operator for approximation and integration in econometrics, to extend the Metropolis-Hastings algorithm, to give an overview, implement and compare some available nonlinear state space filters and develop a software for the estimation of a general nonlinear model class.

The asymptotic convergence analysis suggests that Monte Carlo integration does not exploit smoothness of the integrand. If smoothness of a function is defined by the number of finite derivatives then economic models are often characterized by infinitely smooth functions. Quadrature based deterministic approaches have convergence rates which depend on the ratio between dimension and smoothness, see for example in Gerstner and Griebel (2003a). This permits to trade off dimension against smoothness. The Smolyak operator extends approximation and

integration from the univariate setting to the multivariate one. It relieves substantially the curse of dimensionality by modifying the usual Kronecker product. Integration is needed for rational expectations, weighted residuals, state space filters, posterior of structural parameters and marginal likelihoods. An approximation is needed to obtain the optimal policy function. The operator applies for a spectral as well as a finite element approximation and it works behind the scenes as complete polynomials in the perturbation approach. Moreover it could be of interest for nonparametric econometrics where the domain of the involved variables is divided into several subdomains.
The proposed genetic extension of the standard random walk Metropolis-Hastings algorithm allows for an unbiased convergence test and an automated choice of the optimal variance for the candidate parameter vectors. It simplifies a cluster implementation of the estimation and improves the global maximization properties of the algorithm.
Since the likelihood evaluation is the main bottleneck I implement a Smolyak based Gaussian quadrature filter in order to exploit the model smoothness. The Gaussian filter is compared to the most simple but computationally expensive bootstrap filter which is also called sequential importance resampling particle filter.
The implementation for a general model class serves the purpose to solve and estimate a model without adjusting the code. Only the first order conditions and some control parameters of the algorithms have to be provided. The disadvantage of a general approach is that special features of a model cannot be exploited to accelerate the estimation algorithms. On the other hand efforts to improve the algorithmic efficiency can be devoted to one single code.
The simplest alternative estimation technique is GMM. It uses only the moments implied by the first order conditions instead of the whole information available through the solution and the implied dynamics of the model. This information is embedded in the likelihood. It is a complete density of the observed variables and not only some of their moments. Waste of information results in poor small sample properties and is therefore problematic in usual macroeconomic samples of 100 or even less observations. Moreover GMM is not suited for the model selection and a likelihood based approach is indispensable for this purpose. The popularity of GMM is related to its simple implementation where structural parameters can be estimated without solving the model and evaluating its likelihood.
A linear estimator uses a first order Taylor expansion of the first order conditions around the deterministic steady state. The policy function is obtained from the implied quadratic matrix or Riccati equation by a generalized Schur decomposition. The estimator finally uses a likelihood evaluation techniques like the Kalman filter for the involved linear Gaussian state space model. In addition to biased estimates, this procedure imposes certainty equivalence and is thereby likely to miss important quantitative welfare relationships. I use this linear estimation approach to compare it with the nonlinear estimates.

Section 4.2 summarizes the Bayesian framework. Section 4.3 describes the solution method and the Smolyak operator. Section 4.4 presents the general principle of filtering and various approaches to evaluate the likelihood of a nonlinear state space model. Section 4.5 summarizes the Metropolis-Hastings algorithm and the proposed genetic extension. Section 4.7 presents a Monte Carlo simulation study to evaluate the performance of the genetic extension. It reports the estimated parameters and the marginal likelihood of the nonlinear and linearized model. Section 4.8 shortly describes the developed software and section 4.9 concludes.

4.2 Econom(etr)ics

The principle of information accounting in Bayesian econometrics is the same as in positive and normative economics and Geweke (2005) writes:

> *The strategic advantage of Bayesian statistics stems from the fact that its conditioning is driven by the actual availability of information, and its complete integration with the theory of economic behavior under uncertainty, achieved by Friedman and Savage (1948) and Friedman and Savage (1952).*

Hansen and Heckman (1996) write:

> *The rational agents in real business cycle models use this [statistical decision] theory and, as a consequence, are assumed to process information in a highly structured way. Why should the producers of estimates for the real business cycle models act differently?*

A decision unit in the statistical decision theory is defined by the utility function, variable of interest, policy, models and prior information $\mathcal{D} = \{ U, \boldsymbol{\omega}, \boldsymbol{x}, M, p \}$. Utility $U(\boldsymbol{\omega}, \boldsymbol{x})$ is obtained from the variable of interest $\boldsymbol{\omega}$ and the policy $\boldsymbol{x}$.[1] A decision is usually assisted by alternative models $M = \{M_1, ...M_m\}$. A model M_i is represented by the likelihood or density of observables given unobservables $p(\boldsymbol{y} \,|\, \boldsymbol{\theta}_{M_i}, M_i)$. Observables $\boldsymbol{y}$ and their realizations $\boldsymbol{y}^0$ are explained by unobservables $\boldsymbol{\theta}_{M_i}$. Unobservables are structural parameters $\boldsymbol{\theta}^p_{M_i} \subset \boldsymbol{\theta}_{M_i}$ and states or latent variables $\boldsymbol{s}_{M_i} \subset \boldsymbol{\theta}_{M_i}$. Prior information is encoded in the prior density of unobservables $p(\boldsymbol{\theta}_{M_i} | M_i)$ and the prior probabilities of models $p(M_1), ... p(M_m)$. The variable of interest is any function $p(\boldsymbol{\omega} | \boldsymbol{y}, \boldsymbol{\theta}_{M_i}, M_i)$. Competing models may use different unobservables $\boldsymbol{s}_{M_i} \neq \boldsymbol{s}_{M_j}$ to explain the same observables $\boldsymbol{y}_{M_i} = \boldsymbol{y}_{M_j}$. Progress in economics often amounts to unify contradicting theories in a more general one with less unobservables $\boldsymbol{s}_{M_i} \cup \boldsymbol{s}_{M_j} \supset \boldsymbol{s}_{M_k}$. Bayesian statistics unifies the parameter estimation and model selection and incorporates the parameter and model uncertainty in a coherent decision framework.

[1] Bold symbols are vector or matrix valued. $p(y|x)$ is a conditional density and may also represent a deterministic function $y = f(x)$.

Classical estimates are based on the likelihood $\mathcal{L}(\boldsymbol{\theta}_{M_i}; \boldsymbol{y}) \equiv p(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_i}, M_i)$. Parameter estimates are obtained from the likelihood evaluated at the observed sample $\mathcal{L}^0(\boldsymbol{\theta}_{M_i}) \equiv \mathcal{L}(\boldsymbol{\theta}_{M_i}; \boldsymbol{y}^0)$ for example as the parameter vector maximizing the likelihood $\hat{\boldsymbol{\theta}}_{M_i} = \text{argmax } \mathcal{L}^0(\boldsymbol{\theta}_M)$. Often ad hoc criteria are used to obtain estimators for example when minimizing the distance between theoretical and empirical impulse response functions $\hat{\boldsymbol{\theta}}_{M_i} = \text{argmin } d(\boldsymbol{\theta}_{M_i}, \boldsymbol{y}^0)$. Classical statistics bases inference on the likelihood, a density conditional on unobservables, by comparing it to data. Bayesian statistics is based on a density conditioned on observables and thereby incorporate the uncertainty about unobservables in the decision. A complete Bayesian model M specification with the variables of interest, density of observables and prior density provides the joint density

$$p(\boldsymbol{\omega}, \boldsymbol{y}, \boldsymbol{\theta}_M \mid M) = p(\boldsymbol{\omega} \mid \boldsymbol{y}, \boldsymbol{\theta}_M, M)\, p(\boldsymbol{y} \mid \boldsymbol{\theta}_M, M)\, p(\boldsymbol{\theta}_M \mid M).$$

The posterior of unobservables originates from its prior transformed by the likelihood and the marginal likelihood according to the Bayes formula

$$p(\boldsymbol{\theta}_{M_i} \mid \boldsymbol{y}^0, M_i) = \frac{p(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_i}, M_i)\, p(\boldsymbol{\theta}_{M_i} \mid M_i)}{p(\boldsymbol{y}^0 \mid M_i)}.$$

Evidence of a model updates information about unobservables normalized by the marginal likelihood. The marginal likelihood of a model is given by

$$p(\boldsymbol{y}^0 \mid M_i) = \int p(\boldsymbol{y}^0 \mid \boldsymbol{\theta}_{M_i}, M_i)\, p(\boldsymbol{\theta}_{M_i} \mid M_i)\, d\boldsymbol{\theta}_{M_i}.$$

It allows data to assign probabilities to models $M = \{M_1, ..., M_m\}$

$$p(M_i \mid \boldsymbol{y}^0, M) = \frac{p(\boldsymbol{y}^0 \mid M_i)\, p(M_i)}{p(\boldsymbol{y}^0 \mid M)} = \frac{p(\boldsymbol{y}^0 \mid M_i)\, p(M_i)}{\sum_{j=1}^{m} p(\boldsymbol{y}^0 \mid M_j)\, p(M_j)}.$$

The ratio of two marginal likelihoods is the Bayes factor. It transforms the prior information about the estimated models into the posterior odds ratio

$$\frac{p(M_i \mid \boldsymbol{y}^0)}{p(M_j \mid \boldsymbol{y}^0)} = \frac{p(M_i)}{p(M_j)} \frac{p(\boldsymbol{y}^0 \mid M_i)}{p(\boldsymbol{y}^0 \mid M_j)}.$$

The variable of interest, weighted by the posterior density of unobservables

$$p(\boldsymbol{\omega} \mid \boldsymbol{y}^0, M_i) = \int p(\boldsymbol{\omega} \mid \boldsymbol{y}^0, \boldsymbol{\theta}_{M_i}, M_i)\, p(\boldsymbol{\theta}_{M_i} \mid \boldsymbol{y}^0, M_i)\, d\boldsymbol{\theta}_{M_i}$$

and the model density

$$p(\boldsymbol{\omega} \mid \boldsymbol{y}^0, M) = \sum_{i=1}^{m} p(\boldsymbol{\omega} \mid \boldsymbol{y}^0, M_i)\, p(M_i \mid \boldsymbol{y}^0, M),$$

imply the expected utility

$$\mathrm{E}\left(U\left(\boldsymbol{\omega},\,\boldsymbol{x}\right)|\,\boldsymbol{y}^{0},\,M\right)=\int U\left(\boldsymbol{\omega},\,\boldsymbol{x}\right)p\left(\boldsymbol{\omega}\,|\,\boldsymbol{y}^{0},\,M\right)d\boldsymbol{\omega}.$$

In economics the basic decision unit is a household with prices and consumption as a variable of interest. In a decision with a researcher and institutions as decision units $\mathcal{D}=\{\mathcal{D}^{R},\,\mathcal{D}^{I}\}$ the involved priors and models can in general be different. Muth (1961) pointed towards the need to base the explanation of the behavior of economic institutions on the same model as the researcher and used for this purpose rational expectations. By that he overshot the mark since researchers only search for the appropriate model without knowing it for sure. The assumption of different models for the researcher and the observed opens the door to a robust control analysis and provides the technical means to complete the underlying idea of rational expectations by taking into account the differences in the informational endowment of the researcher and the observed.
The variables of interest in economics are causal relationships between observables and unobservables. Models, parameter estimates and specification test statistics like model selection criteria are the available policies. Econometrics is based on conditional independence assumptions. Their relations to causality are still insufficiently understood and therefore not formalized. The reason lies probably in enduring problems of a satisfactory philosophical definition of causality. The symptom is that the most sophisticated instruments in everyday econometrics are exogeneity and Granger causality. Artificial intelligence research depends on a detailed formalization for causality inference in order to transform computers into more than fast calculating silicon. It goes well beyond exogeneity and Granger causality analysis. The Bayesian information accounting allows to combine the conditional independence and causal assumptions in Bayes networks similar to structural equations in traditional econometrics. Graph theory and further formalisms allow to analyze the networks for causality inferences. Many concepts in econometrics like Granger causality, instrument variables, identification, Lucas' critique, Occam's razor, spurious regressions or counterfactual policies are related in a formal framework to derive probabilities of causation. It transcends the often enough hand waving treatment of these difficult concepts. Pearl (2000) introduces to this most fascinating reasoning which might shape the future developments in econometrics towards a more formal causality inference. Together with the statistical decision framework of a Bayesian approach the production function of econometrics and a microfoundation of econometrics might emerge.
The eight formulas summarizing the statistical decision theory are only apparently simple. The computational challenges of a general nonlinear Bayesian likelihood approach are substantial. Many high dimensional implicit functions and integrals have to be approximated to obtain one single likelihood evaluation at some structural parameters. A practical computational problem is the curse of dimensionality common to many numerical algorithms. It implies that the com-

putational effort rises exponentially with the dimension of the problem. This is very relevant for example in international macroeconomics where theoretical models go well beyond the standard model with two countries each represented by one agent and one technology or portfolio decisions where many asset prices have to be modelled.
The curse of dimensionality appears if a univariate approximation and integration operator is extended from the univariate to the multivariate case by the Kronecker product. One continuous random shock can be approximated for example by a discrete variable with five possible realizations. For a model with 30 shocks, $5^{30} > 9 \times 10^{20}$ integrand evaluations are needed to approximate the rational expectation for one variable. This huge number of function evaluations is also needed for the approximation of a policy of 30 variables by a forth order polynomial. In order to solve the model it has to be done some 100 times to converge within a root finding algorithm at an implicitly defined policy function. For a given vector of structural parameters the policy function implies a likelihood. It is the sum of period contributions and each is a multivariate nonlinear integral. A maximum likelihood estimation will have to evaluate the likelihood at some 100 parameter vectors in addition to many evaluations needed to obtain numerical first derivatives. For a Bayesian estimate of the posterior density of structural parameters some 10,000 likelihood evaluations are needed within a simulation based density approximation. Computational speed considerations are obviously an important topic and gains allow to go beyond highly stylized models.
An intuition of the Smolyak operator can be gained by looking at a multidimensional Taylor expansion. It does not use Kronecker products of univariate polynomials to obtain multivariate polynomials. Instead they are constructed as products of univariate polynomials in such a way that the sum of the exponents of the univariate factors is beneath a certain number which characterizes the approximation accuracy. These polynomials are already known as complete polynomials, see for example in Judd (1998), but it was not clear how to obtain the points where the approximated function has to be evaluated. Since complete polynomials are a special case of the Smolyak operator it can be used to operate on the points for function evaluation in one dimension in order to construct the points in many dimensions.
Another interesting feature of the operator is its hierarchical structure which offers an estimator of the approximation accuracy so that an accuracy depending stopping criterion can be applied instead of an à priori given degree of approximation. This feature can be implemented easily for the integration operator by adding more evaluation points and reweighting the integrand at the old points. It is potentially useful in the estimation process where the model is solved at many different parameter vectors. Some of them may require only a low degree of approximation for a sufficient accuracy while others may imply a higher degree.

The general model representation is the one used by Miranda and Fackler (2002)

$$\begin{aligned} \mathbf{0} &= \boldsymbol{f}(\boldsymbol{s}_t, \boldsymbol{x}_t, E_t\boldsymbol{h}(\boldsymbol{s}_t, \boldsymbol{x}_t, \boldsymbol{e}_{t+1}, \boldsymbol{s}_{t+1}, \boldsymbol{x}_{t+1})) \\ \boldsymbol{s}_{t+1} &= \boldsymbol{g}(\boldsymbol{s}_t, \boldsymbol{x}_t, \boldsymbol{e}_{t+1}). \end{aligned}$$

The variables have to be grouped into state vectors $\boldsymbol{s}$ and policy vectors $\boldsymbol{x}$. The structural state shocks $\boldsymbol{e}$ are in general not Gaussian and independently distributed. $\boldsymbol{f}$ are dynamic first order optimality conditions and $\boldsymbol{g}$ are specified state transition functions. Economic models are functional equations and solutions are policy functions in terms of states $\boldsymbol{x}_t = \boldsymbol{x}^*(\boldsymbol{s}_t)$. The policy functions are defined only implicitly in the first order conditions and can usually not be derived in closed form. Therefore numerical approximation methods have to be used.
The general approach to solve functional equations is to restrict the search for the solution function in the infinite dimensional space of all functions to a lower dimensional space. Functions within this smaller space are represented by a vector of parameters. They are calculated to fit a linear combination of simple basis functions to policy function values at a certain grid. To identify these parameters the function which is being approximated needs to be evaluated at some points. This is a similar problem as in basic econometrics, where these points and the function evaluations are given by data for the independent and dependent variables, the functional form is linear and the criterion to obtain the parameters is least squares. In numerical functional problems we can choose the points where to evaluate the function, the functional form fitted to data and the criterion to identify the parameters of the approximation.
In economic models the policy functions are defined only implicitly and the evaluation of the policy functions cannot be done directly. For a given vector of structural parameters the first order conditions are evaluated at a grid of states and policies. The policy values are changed by function iteration or some general nonlinear root finding algorithm until the first order conditions are near the theoretically exact value of zero.
In an estimation algorithm tens of thousands of nonlinear solutions have to be calculated. Beside speed considerations there are two important topics within this process. One is about finding sensible start values for the policy functions and the other dwells on a robust way to arrive at a solution. In calibration applications where the model is solved for some few parameter vectors this can be handled by searching manually for start values which converge to a plausible solution checked by suitable approximation error estimates. This manual procedure is not practicable in econometric applications and I program a solver relying on numerical linearization in order to generate start values for the nonlinear solver. The likelihood of the linearized model is moreover evaluated by the Kalman filter and provides a check whether an alternative nonlinear solution and estimation improves the fit. The relative fit of these models is examined by the marginal likelihood which is suited as a model selection criterion for nonnested models.

After the solution of a model is obtained, its reduced form can be given in a nonlinear state space form in order to evaluate the likelihood. It is a general time series model and its most attractive features are the distinction between observed and unobserved or latent variables and a recursive updating of information based on the Bayes' formula and the conditional independence property of the involved Markovian process. This allows an integrated approach to nonstationary and cointegrated processes, stochastic trends like the one assumed by the Hodrick-Prescott filter, missing observations, regime switching, GARCH process or in general time varying parameters, learning processes, data revision and robust control applications. In the general nonlinear state space model the distinction between a parameter and a state becomes blurred. Both are unobservables and a parameter can be assumed to be time varying described by some process. By that way the original fixed parameter becomes an unobserved state and the process describing the time dependency is itself described by some fixed parameters.
The solution of the model $\boldsymbol{x}_t = \boldsymbol{x}^*(\boldsymbol{s}_t)$ implies dynamics given by the state transition equation

$$\boldsymbol{s}_{t+1} = \boldsymbol{g}(\boldsymbol{s}_t, \boldsymbol{x}^*(\boldsymbol{s}_t), \boldsymbol{e}_{t+1}) = \boldsymbol{g}^*(\boldsymbol{s}_t, \boldsymbol{e}_{t+1})$$

which together with a measurement equation

$$\boldsymbol{y}_t = \boldsymbol{m}(\boldsymbol{s}_t, \boldsymbol{\epsilon}_t)$$

forms the state space representation of the model. The measurement equation links the unobservable states to observables $\boldsymbol{y}$ with measurement errors $\boldsymbol{\epsilon}$.
The likelihood of a state space model can be obtained as a by-product when deriving the posterior densities of the unobserved states recursively for each period. The densities involved in updating are classified according to the conditioning set. $p(\boldsymbol{s}_\tau \mid \boldsymbol{y}_{1:t})$ is the prior or prediction density for $\tau > t$, filter or posterior density for $\tau = t$ and smoothing density for $\tau < t$. The prior is updated to become the posterior after new data is incorporated and the smoothing density is obtained after the state densities are conditioned on all available data.
The posteriors can be estimated by a bootstrap filter which allows for any shock distribution and nonlinear functions $\boldsymbol{g}$ and $\boldsymbol{m}$. The bootstrap filter approximates the posterior by simulating complete densities. It is a good accuracy benchmark but computationally involved due to its general nature. The likelihood is derived by multivariate integral equations and the particle filter is a Monte Carlo integration method. It has to be combined with a random number generator since the difficulty is to derive draws from a density without a closed form expression. If functions $\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{m}$ and $\boldsymbol{x}^*$ are smooth, the computational efficiency can be gained by exploiting this property.
An overview of some available approaches to nonlinear filtering is given. The general filtering problem is formulated and alternative solutions are discussed. The purpose of this part is to work out where integral equations arise and deterministic integration by the Smolyak operator is possible. The most common

deterministic approach is the unscented filter developed by Julier and Uhlmann (1997) and refined in Julier and Uhlmann (2002). It uses a low order approximation to the integrals involved in order to calculate two moments of a nonlinear transformation of a Gaussian random variable. Structural econometric models result in state transition functions $\boldsymbol{g}$ where the policy solution is plugged in. These policy functions are often of a higher polynomial degree than the unscented filter is derived for. Filtering needs to integrate the state transition functions and the unscented integration is probably not accurate enough for structural econometric applications.
The unscented Kalman and the bootstrap filter can be classified as two possible extreme approaches. The former uses only two moments of a low order approximation for filtering whereas the latter uses complete simulated densities. In general a filter with a flexible approximation degree is desirable since models can exhibit different degrees of nonlinearity and smoothness. A quadrature based nonlinear filter with different degrees of approximation is therefore a compromise between these two extremes and thereby allows to trade off the model dimension against smoothness.
Once a likelihood evaluation procedure is established it can be used for a parameter estimation. In this work I estimate the posterior density of structural parameters by a Metropolis-Hastings algorithm. The parameter posterior is proportional to the likelihood times prior. A posterior density without a closed form can be approximated by a histogram of a sample of random draws. The Metropolis-Hastings algorithm offers a general method to draw random numbers from a density. The only prerequisite is that the density can be evaluated at any point in its domain. The algorithm travels through the feasible parameter space and a simple criterium decides whether the next candidate vector is accepted as a new realization from the posterior or whether the last vector is used again in order to generate another candidate.
Fernádez-Villaverde and Rubio-Ramírez (2004b) use a Metropolis-Hastings algorithm with one sequence of structural parameters generated by a random walk through the parameter space. It is a Markov Chain because only the last parameter vector determines how the next proposal vector is generated. The variances of the random walk shocks have to be tuned before estimation. The recommendation is to choose them in such a way that the sequence exhibits the optimal acceptance ratio around 0.3. The acceptance ratio measures how many candidate vectors are accepted as a fraction of all draws. A higher variance lowers the ratio and vice versa. It is quite difficult to tune these variances in a satisfactory way for more than one parameter.
Compared to a likelihood approach within a derivative based maximization routine there are several advantages of the Metropolis-Hastings algorithm. The derivative based maximization may not converge and end up in local maxima. Either different start values have to be tried manually or a global approach like homotopy or genetic methods have to be implemented. The Metropolis-Hastings

algorithm is easy to implement. It is also a simple global genetic hill climber. Another advantage is that estimated uncertainty about the parameters does not rely on the asymptotic theory and sample size dependent uncertainty estimates are obtained. The uncertainty about parameter estimates is moreover properly incorporated in the decision. as opposed to classical econometrics where some bootstrap approaches have to be used. This is important for example in forecasts or for calibrated parameters. The common calibration approaches to empirical model evaluation can also be improved by incorporating the parameter uncertainty in the prior density. Finally the parameter posterior sample generated with this algorithm provides the starting point for the density estimate of any variables of interest. It can be any function of observables and unobservables like specification test statistics, welfare measures for the policy recommendation and evaluation.

I present a genetic extension of the basic random walk Metropolis-Hastings algorithm to solve its covariance choice problem in an automatic fashion. The innovation is to combine several parallel sequences and to allow each sequence to be driven by random walk innovations and a mixture of parameter vectors from the parallel sequences. This genetic extension reduces the number of free parameters to be tuned to two, so that the recommended acceptance ratio is obtained with less and less expensive trial sequences. The intuition of the algorithm extension can be related to the way a standard diagnostic test for convergence of the algorithm is obtained. Between and within moments should converge to assure that several sequences are drawn from the same density. Since the optimal random walk shock variance is the variance of the unknown target density one might think that the genetic extension allows to estimate the appropriate covariance matrix from parallel sequences. The parallel estimate cannot be obtained from one sequence since parallel sequences start from different parameter vectors. The global maximization properties of a one sequence algorithm are improved and a global likelihood maximizer can be used as an alternative to the posterior density estimation. I used it as the first step to find the modes of the estimated density before the sampling starts. Moreover parallel sequence allow for an unbiased convergence test. The diagnostic test is needed to assure that the algorithm has converged and the generated parameter vectors represent draws from the posterior density independent of start values.

Two other properties within the Metropolis-Hastings algorithm are exploited to accelerate the estimation. Once the sequences have converged subsequent structural parameter vector draws differ only by small amounts and the policy function at the previous vector can be used as a start value. This feature combined with a derivative based root finding algorithm reduces the computations for solving the model to only a small fraction of the overall computations. The main work is the likelihood evaluation for a given structural parameter vector and its associated policy function.

The second acceleration results from the use of constant approximation bounds for the solution algorithm as well as the estimation process. This implies a constant inverse matrix for calculating the Chebyshev coefficients so that fast Fourier transformation and Smolyak specific implementations decrease only the fixed costs of the matrix inversion.
The possibility of the Bayesian framework to select models regarding their ability to fit the data also seems interesting enough. The standard procedure is an informal check how the model fits some stylized facts, a likelihood ratio test or a information criterion like BIC. The drawback of a likelihood ratio test is that the models have to be nested in such a way that one emerges from the other by a parameter restriction. Moreover these tests rely on the fiction of a true model and their small sample properties are not clear either. The Bayesian information criterion is in fact an approximation to the Bayes factor used here and the posterior odds ratio is the proper decision theoretical model selection criterion. As in Fernández-Villaverde and Rubio-Ramírez (2004a) it can be estimated once the Metropolis-Hastings algorithm has converged and produced a sample from the posterior density of structural parameters. With the construct of the variable of interest at hand, out-of-sample forecast performance of a model can also be measured. Such a forecast for future periods $\boldsymbol{y}_t$ with $t > T$ is just another unobservable to be simulated by means of the posterior density.

4.3 Optimal Policy

The general model representation in this work is the one used in Miranda and Fackler (2002). It is summarized in table 4.1. The vector valued function $\boldsymbol{f}$ represents the necessary first order conditions of a dynamic optimization. Distorted, decentralized and social planner equilibrium models can be written in this form and encompass the majority of currently used macroeconomic models. The first order conditions for a social planner equilibrium can be derived by the Bellman functional. The vector valued function $\boldsymbol{g}$ represents the state transition laws of the economy which are shock processes and other state transitions, like the productivity and capital transition in the example model. $\boldsymbol{h}$ represents the forward looking part of the model. The variables have to be classified as state $\boldsymbol{s}$ or policy $\boldsymbol{x}$ (also called action, response or decision variable). The difference is that the transition laws for the state variables are given as part of the model specification whereas the optimal reaction of the policy variables with respect to the states are the object of interest when solving the model.
Mathematically the equation is a functional, like integral or differential equations, and the solution is the policy function. It prescribes how policy should be conducted depending on the state variables so that it satisfies the first order conditions and is therefore dynamically optimal. Since these conditions are usually

Table 4.1: General Macroeconomic Model

Model			
$\boldsymbol{f}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{z})$	$=$	$\mathbf{0}$	
$\boldsymbol{z}$	$=$	$\mathrm{E}_{e'}\boldsymbol{h}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{e}',\boldsymbol{s}',\boldsymbol{x}')$	
$\boldsymbol{s}'$	$=$	$\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{e}')$	
Variables			
$\boldsymbol{s}\in S$	$\subseteq$	$\mathbb{R}^{d_s}$	state variables
$\boldsymbol{x}\in[\boldsymbol{a}(\boldsymbol{s}),\boldsymbol{b}(\boldsymbol{s})]$	$\subseteq$	$\mathbb{R}^{d_x}$	policy variables
$\boldsymbol{z}\in$		$\mathbb{R}^{d_z}$	expectational variables
$\boldsymbol{e}\in$		$\mathbb{R}^{d_e}$	stochastic shocks
Model Functions			
$\boldsymbol{f}:\mathbb{R}^{d_s+d_x+d_z}$	$\to$	$\mathbb{R}^{d_x}$	equilibrium conditions
$\boldsymbol{h}:\mathbb{R}^{d_s+d_x+d_e+d_s+d_x}$	$\to$	$\mathbb{R}^{d_z}$	expectation functions
$\boldsymbol{g}:\mathbb{R}^{d_s+d_x+d_e}$	$\to$	$\mathbb{R}^{d_s}$	state transitions
$\boldsymbol{x}^e:\mathbb{R}^{d_s+d_z}$	$\to$	$\mathbb{R}^{d_x}$	expectational solution
Approximated Functions			
$\boldsymbol{f}(\boldsymbol{s},\boldsymbol{x},\mathrm{E}_{e'}\boldsymbol{h}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{e}',\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x},\boldsymbol{e}'),\boldsymbol{x}'))=0$			
$\boldsymbol{x}^*:\mathbb{R}^{d_s}$	$\to$	$\mathbb{R}^{d_x}$	policy functions
$\boldsymbol{f}(\boldsymbol{s},\boldsymbol{x}^*(\boldsymbol{s}),\sum_j w_j\boldsymbol{h}(\boldsymbol{s},\boldsymbol{x}^*(\boldsymbol{s}),e'_j,\boldsymbol{s}'_j,\boldsymbol{x}^*(\boldsymbol{s}'_j))=0$			
$\boldsymbol{s}'_j=\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x}^*(\boldsymbol{s}),\boldsymbol{e}'_j)$			
$\boldsymbol{z}^*:\mathbb{R}^{d_s}$	$\to$	$\mathbb{R}^{d_z}$	expectation functions
$\boldsymbol{f}(\boldsymbol{s},\boldsymbol{x}^e(\boldsymbol{s},\boldsymbol{z}^*(\boldsymbol{s})),\sum_j w_j\boldsymbol{h}(\boldsymbol{s},\boldsymbol{x}^e(\boldsymbol{s},\boldsymbol{z}^*(\boldsymbol{s})),\boldsymbol{e}'_j,\boldsymbol{s}'_j,\boldsymbol{x}^e(\boldsymbol{s}'_j,\boldsymbol{z}^*(\boldsymbol{s}'_j)))=\mathbf{0}$			
$\boldsymbol{s}'_j=\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x}^e(\boldsymbol{s},\boldsymbol{z}^*(\boldsymbol{s})),\boldsymbol{e}'_j)$			
time is denoted by $v\equiv v_t$, $v'\equiv v_{t+1}$ for $v=\boldsymbol{s},\boldsymbol{x},\boldsymbol{z},\boldsymbol{e}$			

nonlinear and contain an expectation operator they cannot be explicitly solved for the policy and approximation methods are to be used.

There exist at least two possible approximation strategies within the general model. The first approximates the policy functions $\boldsymbol{x}^*(\boldsymbol{s})$ and the second the expectations $\boldsymbol{z}^*(\boldsymbol{s})$. Both approaches are equivalent since $\boldsymbol{x}^*(\boldsymbol{s})$ can be recovered from $\boldsymbol{h}(\boldsymbol{s},\boldsymbol{x}(\boldsymbol{s}))$ and vice versa. I implemented the policy function approximation. The notion $\boldsymbol{x}\in[\boldsymbol{a}(\boldsymbol{s}),\boldsymbol{b}(\boldsymbol{s})]$ allows for state dependent inequality constraints in the model formulation. The resulting Kuhn-Tucker inequality conditions can be transformed into nonlinear max/min equations with smoothed kinks in such a way that they can be solved by a usual root finder. Inequality restrictions are for example liquidity constraints in OLG models saying that young generations cannot borrow against future income prospects. This will result in kinks in the policy functions. Such functions cannot be approximated very well with global polynomials and finite elements should be used. Both are discussed in section

4.3.3 but in the current software only global polynomials are implemented. However, a finite element approximation is a rather simple extension of the current implementation.

4.3.1 Utility Maximization

The example model is the canonical dynamic consumption and labor choice decision. It is described by the following constrained maximization

$$\max_{\{c_t,l_t\}_{t=0}^{\infty}} U = \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t,l_t) = \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{(c_t^\theta (1-l_t)^{1-\theta})^{1-\tau}}{1-\tau} \tag{4.1}$$

subject to

$$y_t = c_t + i_t \tag{4.2}$$

$$y_t = e^{a_t} k_t^\alpha l_t^{1-\alpha} \tag{4.3}$$

$$k_{t+1} = i_t + (1-\delta)k_t \tag{4.4}$$

$$a_{t+1} = \rho a_t + e_{t+1} \quad \text{where } e_t \overset{iid}{\sim} \mathcal{N}(0,\sigma_e). \tag{4.5}$$

The variable of interest in the decision of equation (4.1) is the maximum of the present value of the utility from consumption c_t and leisure $1-l_t$ policy. Leisure is given by the time remaining after labor l_t is subtracted from the available time normed to 1. The structural parameters are the discount factor β, the elasticity of intertemporal substitution τ, the labor supply determinant θ, the technical substitution α, the depreciation rate δ and the autocorrelation coefficient ρ of the productivity process a_t. E_0 are expectations conditional on available information at time 0. Equation (4.2) is the budget constraint, restricting output y_t to be either consumed c_t or invested i_t. Equation (4.3) describes the production technology with capital k_t, labor l_t and productivity a_t. Equation (4.4) is the law of motion of capital and equation (4.5) describes the process of productivity with productivity shock e_t.

4.3.2 First Order Conditions

The model has no distortions. Hence both welfare theorems apply and the necessary first order conditions can be derived by the Bellman functional as an optimization problem of a social planner. To facilitate the derivation, I substitute two out of three constraints by plugging (4.3) and (4.4) into (4.2) and eliminate y_t and i_t. The Bellman equation specifies the value function

$$\mathrm{V}(k_t,a_t) = \max_{c_t,\,l_t,\,k_{t+1}} u(c_t,l_t) + \beta\,\mathrm{E}_t\mathrm{V}(k_{t+1},a_{t+1}) + \lambda(k_{t+1} - e^{a_t}k_t^\alpha l_t^{1-\alpha} + c_t - (1-\delta)k_t).$$

The first derivatives with respect to the decision variables are

$$\frac{\partial \mathrm{V}(k_t, a_t)}{\partial c_t} = \theta \frac{(c_t^\theta (1-l_t)^{1-\theta})^{1-\tau}}{c_t} + \lambda \stackrel{!}{=} 0 \tag{4.6}$$

$$\frac{\partial \mathrm{V}(k_t, a_t)}{\partial l_t} = -(1-\theta) \frac{(c_t^\theta (1-l_t)^{1-\theta})^{1-\tau}}{1-l_t} - \lambda(1-\alpha) \frac{e^{a_t} k_t^\alpha l_t^{1-\alpha}}{l_t} \stackrel{!}{=} 0 \tag{4.7}$$

$$\frac{\partial \mathrm{V}(k_t, a_t)}{\partial k_{t+1}} = \beta \mathrm{E}_t \mathrm{V}_{k_{t+1}}(k_{t+1}, a_{t+1}) + \lambda \stackrel{!}{=} 0. \tag{4.8}$$

The derivative $\mathrm{V}_{k_t+1}(k_{t+1}, a_{t+1})$ has to be substituted since the value function is unknown. The envelope theorem allows to omit the derivatives of the unknown policy function with respect to the states since at optimum they are zero. The derivative of the value function is therefore

$$\begin{aligned} \mathrm{V}_{k_t}(k_t, a_t) &= -\lambda(1 - \delta + \alpha \frac{e^{a_t} k_t^\alpha l_t^{1-\alpha}}{k_t}) \\ &= \theta \frac{(c_t^\theta (1-l_t)^{1-\theta})^{1-\tau}}{c_t} \left(1 - \delta + \alpha \frac{e^{a_t} k_t^\alpha l_t^{1-\alpha}}{k_t}\right) \end{aligned}$$

where λ is substituted by equation (4.6). The derivative can be forwarded one period and used to equate (4.8) and (4.6) through λ. This gives the first out of two necessary first order condition. It is an Euler equation which characterizes the optimal intertemporal consumption and labor policy

$$\frac{(c_t^\theta (1-l_t)^{1-\theta})^{1-\tau}}{c_t} - \beta \mathrm{E}_t \left(\frac{(c_{t+1}^\theta (1-l_{t+1})^{1-\theta})^{1-\tau}}{c_{t+1}} \left(1 - \delta + \alpha \frac{e^{a_{t+1}} k_{t+1}^\alpha l_{t+1}^{1-\alpha}}{k_{t+1}}\right)\right) = 0. \tag{4.9}$$

An optimal decision balances present and future marginal rewards. The present reward is simply the present marginal utility whereas the future reward is the future marginal utility times the gross return which depends on the discount factor, depreciation and the marginal productivity of capital. Using (4.6) and (4.7) to eliminate λ gives us the second necessary first order condition

$$\frac{1-\theta}{\theta} \frac{c_t}{1-l_t} - (1-\alpha) \frac{e^{a_t} k_t^\alpha l_t^{1-\alpha}}{l_t} = 0. \tag{4.10}$$

It is a static equation defining an optimal intratemporal trade off between consumption and labor.

The numerical procedures can be used to find two policy functions for labor and consumption implicitly defined in the last two equations. This is inefficient since the static optimality condition can be solved explicitly for consumption in terms of labor. The solution can be plugged into the Euler equation turning it into an equation describing the optimal labor supply. Once the optimal labor decision is approximated the optimal consumption can be recovered from the static equation.

In order to set out the problem in a general multivariate setting I do not use this analytical solution in the discussion but only in the numerical routines.
Now the system can be mapped into the general form of table 4.1. The state vector is $\boldsymbol{s}_t \equiv \{k_t, a_t\}$ with $d_s = 2$, the policy variables are $\boldsymbol{x}_t \equiv \{c_t, l_t\}$ with $d_x = 2$ and the productivity shock in equation (4.5) corresponds to the shock in the general model form, $\boldsymbol{e}_t \equiv e_t$ with $d_e = 1$. There is only one expected variable $\boldsymbol{z}_t$ with $d_z = 1$ and its functional form $\boldsymbol{h}$ is given by the argument of the expectation operator in equation (4.9). Equations (4.9) and (4.10) correspond to the vector valued function $\boldsymbol{f}$ with $d_x = 2$. The capital and productivity transition equations

$$\begin{aligned} k_{t+1} &= e^{a_t} k_t^{\alpha} l_t^{1-\alpha} - c_t + (1-\delta)k_t \\ a_{t+1} &= \rho a_t + \epsilon_t \end{aligned}$$

correspond to the vector valued state transition function $\boldsymbol{g}$ with $d_s = 2$. The model solution $\boldsymbol{x}^*(\boldsymbol{s})$ implies the state transition

$$\boldsymbol{s}_{t+1} = \boldsymbol{g}(\boldsymbol{s}_t, \boldsymbol{x}^*(\boldsymbol{s}_t), \boldsymbol{e}_t) = \boldsymbol{g}^*(\boldsymbol{s}_t, \boldsymbol{e}_t).$$

This is the state transition equation of a state space model and will be discussed in section 4.4 as a tool to evaluate the likelihood of the model.

4.3.3 Approximation

The solution of a functional is a function and resides in an infinite dimensional space. A numerical procedure has to restrict its search to a lower dimensional space where the solution is characterized by a finite vector of parameters $\boldsymbol{c}$. The univariate function approximation is given by a linear combination of basis functions $\psi_i(s)$ of some functional form

$$f(s) \approx \hat{f}(s) = \sum_{i=0}^{n} c_i \, \psi_i(s). \tag{4.11}$$

However, the policy functions $\boldsymbol{x}^*(\boldsymbol{s})$ depend in general on d_s state variables and some further techniques are needed to extend approximation to the multidimensional case. There are three broad numerical methods for function approximation. The perturbation approach is multidimensional by nature. The finite element and spectral approaches are univariate in the beginning and have to be extended to a multivariate operator. The extension methods compared in this work are the Kronecker product and the Smolyak operator and will be discussed in section 4.3.5.
The most popular perturbation method uses Taylor series expansions of the model equations to derive the policy function. Another perturbation method uses Padé series which are ratios of Taylor series. Often a loglinear approximation is used.

Whether this is an improvement can be answered by approximation error estimates and Aruoba, Fernádez-Villaverde, and Rubio-Ramírez (2003) report that for the model used here loglinearization exhibits larger approximation errors. Judd (2002) generalizes this change of variable and shows that the optimal transformations are difficult to derive but dominant can be constructed. The predominant linearization technique is a perturbation method where a first order Taylor expansion around the deterministic steady state is calculated. Perturbation methods rely on the implicit function theorem and use only local information, namely the policy and its first and higher derivatives at the steady state.
One problem is that the approximation is valid only within a ball around the steady state and a radius equal to the distance between the steady state and the next singularity of the function approximated. Valid means that outside this ball the approximation quality is unacceptable. Since the functions we are approximating are defined only implicitly, the singularity can be hardly obtained analytically and an approximation error estimate is indispensable. This is valid for all approximation strategies and one possible error estimate is discussed in section 4.3.9.
Taylor series can be used to gain economic intuition about the dynamic mechanism at work since the functional form in the implied dynamic system $\boldsymbol{s}_{t+1} = \boldsymbol{g}(\boldsymbol{s}_t, \boldsymbol{x}^*(\boldsymbol{s}_t), \boldsymbol{e}_t)$ is known and the parameters are explicitly expressed as functions of the structural parameters. A second order Taylor approximation can be used to avoid imposing the certainty equivalence property on a nonlinear model. Then the policy depends on the state vector and the state shock standard deviations. The perturbation approach is described by Schmitt-Grohé and Uribe (2004) and is now routinely used since software for a general class of models is available. A disadvantage is that higher order approximations are said to require analytical derivatives because numerical derivatives accumulate errors. They should be calculated by symbolic software like Maple or Mathematica. I calculate the linearization by numerical derivatives. The error estimates show that analytical derivatives do hardly improve the approximation quality at the true parameters and do not change the likelihood at all. Another disadvantage is that kinks of the policy function and therefore inequality restrictions cannot be handled.
For the spectral and finite element approximation the borders of the approximation space $S \ni \boldsymbol{s}$ have to be defined. It is not necessary to calculate the steady state for the approximation but it gives an idea about the appropriate approximation interval $S \subset \mathbb{R}^{d_s}$.
The spectral approach use orthogonal polynomials instead of nonorthogonal monomials ($\psi_n(x) = x^n$) of the Taylor expansion. A family of polynomials $\{\psi_n(x)\}$ is mutually orthogonal with respect to a weighting function $w(x)$ if

$$< \psi_n(x), \psi_m(x) > \quad = \int_a^b \psi_n(x)\, \psi_m(x)\, w(x)\, dx = 0 \text{ for } n \neq m.$$

The coefficients of the spectral polynomial are derived from function evaluations at more points in the approximation space then only the steady state as in the perturbation approach. The Chebyshev approximation comes close to the holy grail of the approximation theory: the minimax polynomial. It is the approximating polynomial which has the smallest maximum deviation from the true function. It is closely approximated by Chebyshev polynomials of the first kind used in this work. The disadvantage of this method is that it is not suited to handle kinks in the policy functions. The advantage is that for smooth models it needs less function evaluations compared to the next method. The number of points where the function is evaluated determines the length of the policy defining parameter vector. This vector has to be found within a numerical root finding procedure and shorter vectors results in a much faster algorithm.
The third approximation method is finite elements. It approximates like the spectral method globally but uses local low order polynomials. Finite elements have basis functions which are nonzero only in a small region of the approximation space whereas spectral basis functions are nonzero at almost all points of the approximation space. It effectively divides the approximation space and approximates within these subspaces such that low degree polynomials are sufficiently accurate. The advantage is that it can accurately approximate kinks in policy functions since these local anomalies do not influence the approximation in other subspaces. The disadvantage is that for smooth functions it needs many function evaluations compared to the spectral method.

4.3.3.1 Perturbation

Perturbation is a local function approximation. In a first order perturbation the policy value and its first derivative are used to recover the implicit policy function. Linear rational expectation models result in a quadratic matrix or Riccati equation which can be solved fast and accurate by the generalized Schur or QZ decomposition, see Klein (2000). Why do we arrive at a quadratic equation for the parameters which determines the model solution $\boldsymbol{x}^*(s)$? If the first order conditions $\boldsymbol{f}$, the state transitions $\boldsymbol{g}$ and the expectations function $\boldsymbol{h}$ are linear then the solution $\boldsymbol{x}^*(\boldsymbol{s})$ is linear, too. The dynamic nature of the model introduces a nested application of the solution function $\boldsymbol{x}' = \boldsymbol{x}^*(\boldsymbol{s}') = \boldsymbol{x}^*(\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}^*(\boldsymbol{s}), \boldsymbol{e}))$ within the forward looking function $\boldsymbol{h}$. If for example $x^*(s) = a + bs$ than $x^*(x^*(s)) = a + b(a + bs) = a + ab + b^2 s$, i.e. the solution parameter b appears quadratically and the first order functional has two roots. Economists are usually interested in the one which implies a stable saddle path of the underlying variables in $\boldsymbol{s}_{t+1} = \boldsymbol{g}(\boldsymbol{s}_t, \boldsymbol{x}^*(\boldsymbol{s}_t), \boldsymbol{e}_t)$.
A more accurate solution can be obtained if higher order Taylor polynomials and therefore second and higher order derivatives of the policy function at the steady state are used. The policy at the steady state is different compared to the first order approximation since risk premia are taken into account. This is

achieved by augmenting the state vector on which the policy depends by the standard deviation of the involved shocks. The marginal effort of a second order approximation is low. Beside another derivative of the system it is a simple matrix inversion. Once the saddle path in the first order Taylor approximation is picked out, the solution for the second order approximation is unique.
Now the linearization equations are given for the general model class. First the deterministic steady state $\bar{\boldsymbol{s}}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{e}}$ is needed. The steady state of shocks is their zero expected value $\bar{\boldsymbol{e}} = \mathbf{0}$. For the other variables the deterministic steady state is defined by

$$\begin{aligned} \mathbf{0} &= \boldsymbol{f}(\bar{\boldsymbol{s}}, \bar{\boldsymbol{x}}, \boldsymbol{h}(\bar{\boldsymbol{s}}, \bar{\boldsymbol{x}}, \mathbf{0}, \bar{\boldsymbol{s}}, \bar{\boldsymbol{x}})) \\ \bar{\boldsymbol{s}} &= \boldsymbol{g}(\bar{\boldsymbol{s}}, \bar{\boldsymbol{x}}, \mathbf{0}). \end{aligned}$$

A closed form solution is usually not available and a nonlinear root finder has to solve the equations. The linearized first order equations around the steady state are given by

$$\begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ -\boldsymbol{f}_z \boldsymbol{h}_{s'} & -\boldsymbol{f}_z \boldsymbol{h}_{x'} \end{bmatrix} \begin{bmatrix} d\boldsymbol{s'} \\ d\boldsymbol{x'} \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}_s & \boldsymbol{g}_x \\ \boldsymbol{f}_s + \boldsymbol{f}_z \boldsymbol{h}_s & \boldsymbol{f}_x + \boldsymbol{f}_z \boldsymbol{h}_x \end{bmatrix} \begin{bmatrix} d\boldsymbol{s} \\ d\boldsymbol{x} \end{bmatrix}$$

where $d\boldsymbol{s}$ and $d\boldsymbol{x}$ denote deviations from the steady state and $\boldsymbol{x'}$ and $\boldsymbol{s'}$ are again the next period variables. The dimensions of the identity $\boldsymbol{I}$ and null matrix $\mathbf{0}$ are $d_s \times d_s$ and $d_s \times d_x$ respectively. The subscripted functions $\boldsymbol{f}$, $\boldsymbol{g}$ and $\boldsymbol{h}$ are Jacobians with respect to the variables in the subscript evaluated at the steady state. Therefore the left matrices on both sides of the equation are constants for a given vector of structural parameters. The solution is a $d_x \times d_s$ matrix $\boldsymbol{C}$ with $d\boldsymbol{x} = \boldsymbol{C} d\boldsymbol{s}$. This solution implies the state transition matrix $\boldsymbol{P}$ of size $d_s \times d_s$ in $d\boldsymbol{s'} = \boldsymbol{P} d\boldsymbol{s}$. The matrix $\boldsymbol{C}$ represents the optimal linear policy for a given state vector. The matrix $\boldsymbol{P}$ maps the current state into the expectation of next period state. It is the combined effect of the policy function $\boldsymbol{C}$ in the specified state transition. The resulting matrix system is

$$\begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ -\boldsymbol{f}_z \boldsymbol{h}_{s'} & -\boldsymbol{f}_z \boldsymbol{h}_{x'} \end{bmatrix} \begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{C}\boldsymbol{P} \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}_s & \boldsymbol{g}_x \\ \boldsymbol{f}_s + \boldsymbol{f}_z \boldsymbol{h}_s & \boldsymbol{f}_x + \boldsymbol{f}_z \boldsymbol{h}_x \end{bmatrix} \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{C} \end{bmatrix}.$$

The solutions $\boldsymbol{C}$ and $\boldsymbol{P}$ are obtained by the QZ decomposition of the constant matrices and the system can be written as

$$\boldsymbol{Q} \begin{bmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \mathbf{0} & \boldsymbol{S}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_{11}^H & \boldsymbol{Z}_{21}^H \\ \boldsymbol{Z}_{12}^H & \boldsymbol{Z}_{22}^H \end{bmatrix} \begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{C}\boldsymbol{P} \end{bmatrix} = \boldsymbol{Q} \begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} \\ \mathbf{0} & \boldsymbol{T}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_{11}^H & \boldsymbol{Z}_{21}^H \\ \boldsymbol{Z}_{12}^H & \boldsymbol{Z}_{22}^H \end{bmatrix} \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{C} \end{bmatrix}$$

where $\boldsymbol{Q}$ and $\boldsymbol{Z}$ are unitary, $\boldsymbol{S}$ and $\boldsymbol{T}$ are upper triangular matrices and H denotes the conjugate transpose. To ensure stability for the equilibrium process the decomposition the eigenvalues smaller than 1 in absolute value are placed in the left upper corner, see for example Blanchard and Kahn (1980) or Klein (2000).

The eigenvalues are obtained by element wise ratios of the diagonal elements of $\boldsymbol{T}$ and $\boldsymbol{S}$. Using the properties $\boldsymbol{Z}_{11}^H \boldsymbol{Z}_{11} + \boldsymbol{Z}_{21}^H \boldsymbol{Z}_{21} = \boldsymbol{I}$ and $\boldsymbol{Z}_{12}^H \boldsymbol{Z}_{11} + \boldsymbol{Z}_{22}^H \boldsymbol{Z}_{21} = \boldsymbol{0}$ of unitary matrices the policy function is given by

$$\boldsymbol{C} = \boldsymbol{Z}_{21} \boldsymbol{Z}_{11}^{-1}$$

and the implied state transition by

$$\boldsymbol{P} = \boldsymbol{Z}_{11} \boldsymbol{S}_{11}^{-1} \boldsymbol{T}_{11} \boldsymbol{Z}_{11}^{-1}. \tag{4.12}$$

These solutions can be used to either simulate the system or calculate the likelihood of the model. The matrices $\boldsymbol{C}$ and $\boldsymbol{P}$ are only the slope coefficients of the policy function and equilibrium dynamics. The constants are given by the steady state values and the complete functions are

$$\begin{aligned} \boldsymbol{x}_{t+1}^* &= \bar{\boldsymbol{x}} + \boldsymbol{C}(\boldsymbol{s}_t - \bar{\boldsymbol{s}}) \\ \boldsymbol{s}_{t+1}^* &= \bar{\boldsymbol{s}} + \boldsymbol{P}(\boldsymbol{s}_t - \bar{\boldsymbol{s}}). \end{aligned}$$

The linearized measurement equation of the state space form is

$$\begin{aligned} \boldsymbol{m}(\boldsymbol{s}, \boldsymbol{x}) &\approx \bar{\boldsymbol{m}} + \bar{\boldsymbol{m}}_s(\boldsymbol{s} - \bar{\boldsymbol{s}}) + \bar{\boldsymbol{m}}_x(\boldsymbol{x} - \bar{\boldsymbol{x}}) \\ &\equiv \bar{\boldsymbol{m}} + \bar{\boldsymbol{m}}_s(\boldsymbol{s} - \bar{\boldsymbol{s}}) + \bar{\boldsymbol{m}}_x(\bar{\boldsymbol{x}} + C(\boldsymbol{s} - \bar{\boldsymbol{s}}) - \bar{\boldsymbol{x}}) \\ &= \bar{\boldsymbol{m}} + (\bar{\boldsymbol{m}}_s + \bar{\boldsymbol{m}}_x C)(\boldsymbol{s} - \bar{\boldsymbol{s}}). \end{aligned}$$

where $\bar{\boldsymbol{m}}_{..} \equiv \boldsymbol{m}_{..}(\bar{\boldsymbol{s}}, \bar{\boldsymbol{x}})$ denotes the measurement function and its Jacobians evaluated at the steady state. The slope of the linearized measurement equation is therefore

$$\boldsymbol{M} = \bar{\boldsymbol{m}}_s + \bar{\boldsymbol{m}}_x \boldsymbol{C} \tag{4.13}$$

and the constant term

$$\bar{\boldsymbol{M}} = \bar{\boldsymbol{m}} - \boldsymbol{M}\bar{\boldsymbol{s}} \tag{4.14}$$

where $\bar{\boldsymbol{m}} = \bar{\boldsymbol{y}}$ is the steady state of the observables.

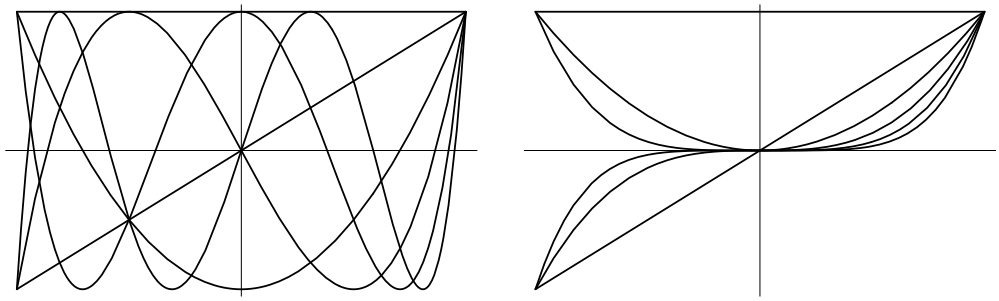
4.3.3.2 Spectral

Spectral methods use basis functions which are nonzero almost everywhere in the approximation hyper cube $S \subset \mathbb{R}^{d_s}$. Basis functions are orthogonal polynomials like Chebyshev polynomials defined by

$$T_0(s) = 1 \qquad T_1(s) = s \qquad T_{i+1}(s) = 2s\,T_i(s) - T_{i-1}(s) \qquad \text{for } i = 1, \ldots$$

The problem of monomials $\psi_i(s) = s^i$ compared to orthogonal polynomials is similar to multicollinearity in regression analysis: information is poorly used if

Figure 4.1: Chebyshev Polynomials and Monomials



regressors are correlated. Monomials are highly correlated and orthogonal polynomials by definition are not. Figure 4.1 shows polynomials of rising degrees. The orthogonal Chebyshev polynomials in the left figure fill the space more uniformly than monomials and result in higher approximation quality. An approximating of a function by the collocation method requires the approximation to be exact at some chosen points. It is compared to other methods in section 4.3.6.
The approximation of function $f(s)$ is given by $\hat{f}(s) = \sum_{i=0}^{n} c_i T_i(s)$. To identify the $n+1$ parameters $\boldsymbol{c} = [c_0\ c_1\ c_2]'$ we have to evaluate the function at $n+1$ points. The $n+1$ collocation conditions exactly determine the solution vector $\boldsymbol{c}$

$$\boldsymbol{T}(\boldsymbol{s})\,\boldsymbol{c} = f(\boldsymbol{s}) \tag{4.15}$$

$$\begin{bmatrix} T_0(s_0) & T_1(s_0) & T_2(s_0) \\ T_0(s_1) & T_1(s_1) & T_2(s_1) \\ T_0(s_2) & T_1(s_2) & T_2(s_2) \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} f(s_0) \\ f(s_1) \\ f(s_2) \end{bmatrix}$$

The basis matrix $\boldsymbol{T}(\boldsymbol{s})$ in equation (4.15) is a constant for given $\boldsymbol{s}$. The orthogonality property of Chebyshev polynomials ensures that this matrix has a low condition number. It can therefore be inverted without numerical problems in a finite precision environment as opposed to a basis matrix of monomials. Left multiplication of the function values with the inverse basis matrix determines the solution parameters $\boldsymbol{c} = \boldsymbol{T}^{-1}(\boldsymbol{s})\, f(\boldsymbol{s})$.
The collocation approach is not interesting for econometrics since $n+1$ data points $f(\boldsymbol{s})$ are translated into $n+1$ parameters $\boldsymbol{c}$ and the statistical goal of dimension reduction is not achieved. An overdetermined system with more points than parameters can be solved for the parameters by least squares.
How should the points $s_0, ..., s_n$ be chosen? The numerical theory says that the optimal points or nodes are the roots of a Chebyshev polynomial. Fortunately there is an explicit solution for Chebyshev polynomial roots given by

$$s_k^T = \cos\left(\frac{\pi(k+0.5)}{n+1}\right) \qquad \text{for } k = 0, ..., n.$$

In the three point example the solutions for $T_3(x) = 4x^3 - 3x = 0$ are $(0, \pm\sqrt{3}/2)$. The solutions are always within the interval $(-1, 1)$ and have to be transformed linearly if another approximation interval is needed. If the approximation interval is $S = [a, b]$ the appropriate nodes are $s_k = a + (b - a)(s_k^T + 1)/2$ for $k = 0, ..., n$. The inverse transformation has to be applied if interpolated function values at other points than s_k are needed. The approximation coefficients are determined with the basis matrix evaluated at the Chebyshev roots $\boldsymbol{s}^T = [s_0^T, ..., s_n^T]$ whereas the function is evaluated at the transformed points $\boldsymbol{s} = [s_0, ..., s_n]$

$$\boldsymbol{c} = \boldsymbol{T}^{-1}(\boldsymbol{s}^T) f(\boldsymbol{s}).$$

For an interpolation aside the nodes at x the inverse linear transformation $x^T = 2(x - a)/(b - a) - 1$ has to be applied where the basis matrix is evaluated at

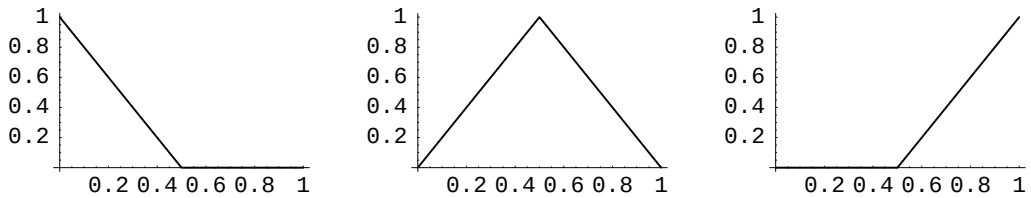
$$\hat{f}(x) = \boldsymbol{T}(x^T)\, \boldsymbol{c}.$$

The complication in economic models is that the policy functions we want to approximate are unknown and we cannot obtain function values at the nodes $f(\boldsymbol{s})$ by simple evaluations. Instead we have to guess start values and improve them according to an accuracy measure. The accuracy is measured by the residuals of the first order conditions when states and policies are plugged in. The residuals should be small since for the exact policy function they are zero for all possible states. The start policy values will usually not provide sufficiently small residuals and some methods are needed to change them in the direction where the residuals approach zero. This problem will be discussed in section 4.3.7. The second problem is that rational expectation models involve an expectation operator in the functional equations and therefore some integrals have to be evaluated before the residual is obtained. This will be discussed in section 4.3.4.

4.3.3.3 Finite Elements

The principle of finite elements is to divide the approximation space into many subspaces where low degree basis functions are used for approximation. Figure

Figure 4.2: Linear Finite Elements



4.2 shows three linear local basis functions known as hat functions in a one-dimensional approximation problem. The approximation space is divided into two

intervals $[0, .5]$ and $[.5, 1]$ by three nodes $\{0, .5, 1\}$. There are always two basis functions which are not zero in each interval. In the first interval the right wing of the hat function with peak at 0 and the left wing of the hat function with peak at 0.5 are not zero. If we want the function to be exactly approximated at the nodes (collocation approach, see section 4.3.6) we need two parameters per interval. These parameters determine the linear combination of the basis functions. They are equal to the function values at the nodes where one basis functions is 1 and the other 0. Therefore the collocation approach with linear splines amounts to solve the nonlinear functional at some nodes. The evaluation of the approximation at other points than the nodes is then a simple linear interpolation in the appropriate interval. As opposed to the spectral approximation the nodes can be chosen freely. In integral equations linear finite elements can be used to approximate the integrand. Once this is done the integral can be easily approximated as the area under the linear functions. Higher order local polynomials should be used if first derivatives have to be smooth since with linear finite elements they have kinks at the nodes.

4.3.4 Integration

Rational expectations in economic models are represented by integration operators. Closed form integrals for nonlinear models are usually unknown and numerical integration is based on deterministic or by random methods.
Monte Carlo integration chooses random points in the interval of integration evaluates the integrand. The average is an estimate of the expected value of the function. Since the integral represents the area under the integrand, the average multiplied by the size of the interval is an estimate of the integral. The usual argument for a Monte Carlo approach is that it is independent of the problem dimension. This is at best misleading since it is the convergence rate which is independent and moreover rather low. The problem with the Monte Carlo approach is that it is often much too general in the sense that smoothness of the integrand is not exploited whereas most economic models are described in terms of smooth functions. Smoothness can be described by bounded mixed derivatives up to order r.
In order to achieve an accuracy ϵ, the convergence for Monte Carlo methods is described by

$$\epsilon(n) = \mathcal{O}(n^{1/2}).$$

It is independent of dimension d and smoothness r, see for example Gerstner and Griebel (2003a). Random numbers in binary computers are not really random but generated by deterministic sequences which mimic some important random number properties. For the purpose of integration better sequences can be gener-

ated which fill the space more uniformly like Niederreiter sequences. They reduce complexity to

$$\epsilon(n) = \mathcal{O}(n^{-1}(\log n)^d).$$

An integration technique which exploits smoothness of the integrand is Gaussian quadrature. Its complexity is

$$\epsilon(n) = \mathcal{O}(n^{-r/d})$$

and the amount of work to achieve a prescribed accuracy grows exponentially with the dimension d. This exponential dependency is known as the curse of dimensionality. The rule of thumb for Kronecker Gaussian quadrature is that it is not feasible for integrals with a higher dimension than four or five.
The complexity of Smolyak Gaussian quadrature is of order

$$\epsilon(n) = \mathcal{O}(n^{-1}(\log n)^{(d-1)(r+1)})$$

and can be expected to outperform Quasi-Monte Carlo methods for smooth functions of order $r > 1$. For very smooth integrands with $r \to \infty$ convergence is even exponential. This method does not suffer from the exponential curse of dimensionality.
Gaussian quadrature was applied in economics for example in Tauchen and Hussey (1991a). Univariate quadrature chooses some deterministic points x_i and weights w_i and takes a weighted mean of the integrand as the integral approximation

$$\int g(x)\,dx \approx \sum_{i=0}^{n-1} w_i g(x_i).$$

The evaluation of the integrand is the most costly part of numerical integration and the amount of work can be measured by the number of nodes n.
A useful starting point to understand how these nodes and weights are derived is to approximate the integrand g by a polynomial

$$g(x) \approx \sum_{i=0}^{n-1} c_i x^i.$$

Quadrature works well for smooth integrands where a small number of nodes and therefore integrand evaluations is sufficient for an accurate approximation. Once the approximating polynomial is obtained, the integral value can be given in closed form

$$\int_a^b g(x)\,dx \approx \int_a^b \sum_{i=0}^{n-1} c_i x^i dx = \sum_{i=0}^{n-1} \int_a^b c_i x^i dx = \sum_{i=0}^{n-1} \left[c_i \frac{x^{i+1}}{i+1} \right]_a^b = \sum_{i=0}^{n-1} c_i \frac{b^{i+1} - a^{i+1}}{i+1}.$$

It is a simple function of the coefficients c_i, the polynomial degree n and the integration bounds a, b. There is no fundamental difference between integration and approximation operators and the Smolyak operator works equally for both. Both operators use function evaluations at some nodes x_i. Whereas function approximation delivers the polynomial coefficients c_i, quadrature derives the weights w_i. The weights are in fact a function of the coefficients of the polynomial which approximates the integrand.
If the function g can be separated into $g(x) = f(x)\,w(x)$ further efficiency can be gained. Gaussian quadrature is the general approach and specialized formulas derive nodes and weights for a given weight function w

$$\int g(x)\,dx = \int f(x)\,w(x)\,dx \approx \sum_{i=1}^{n} w_i\,f(x_i).$$

In general an appropriate family of orthogonal polynomials can be derived for many weight functions which can then be used to approximate the integral, see Press, Flannery, Teukolsky, and Vetterling (1988). For standard weight functions the orthogonal polynomials are known. For example Legendre polynomials, associated with $w(x) = 1$, $\forall x$, are used for an unweighted integration with $g = f$. The estimation of structural parameters as well as rational expectations involve the weighted form of integration where the weight function is a density. In usual economic models Hermite polynomials can be used for the kernel $w(x) = e^{-x^2}$ of a standard normal distribution.
The integral of a density weighted function is the expected value of the transformed random variable

$$\mathrm{E}(f(x)) = \int f(x)\,p\,(x)\,dx.$$

In principle one can integrate the combined function $g(x) = f(x)\,p\,(x)$ by Gauss-Legendre quadrature. Instead of approximating a density function p by a polynomial it is more efficient to represented it by a collection of nodes and associated weights. In the limit of an infinite number of nodes, the weighted sum of integrand values at the nodes is the expected value of interest. Gaussian quadrature of a density weighted function is therefore nothing else than a clever way to discretize the continuous random variable x. The integral is approximated by a weighted sum of integrand evaluations at the nodes. The discrete density approximation is a histogram with n nodes x_i and weights w_i

$$p\,(x) \approx \sum_{i=1}^{n} w_i \delta(x_i - x).$$

The Dirac delta function δ is defined by $\int f(x)\,\delta(x-a)\,dx = f(a)$ and allows to combine the infinitesimal integration operator and its discrete counterpart. The discrete approximation of a density, substituted in the integral functional gives

$$\begin{aligned} \mathrm{E}(f(x)) &\approx \int f(x) \sum_{i=1}^{n} w_i \delta(x_i - x)\,dx \\ &= \sum_{i=1}^{n} w_i f(x_i). \end{aligned}$$

Higher moments of a transformed random variable are expected values of the function raised to higher powers and the same procedure applies to obtain them. This will be needed in the estimation part in section 4.4.
How to choose the nodes and weights for a given density p? As for Chebyshev function approximation the nodes are given by the roots of the involved polynomial. For a normal density the associated Hermite polynomials are given by the recursion

$$H_0(x) = 1 \qquad H_1(x) = 2x \qquad H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x).$$

Unfortunately a closed form solution for larger n does not exist and a nonlinear root finder is needed. Press, Flannery, Teukolsky, and Vetterling (1988) give good start values for different densities such that a simple Newton iteration finds the roots. For a three point approximation of a standard normal density the nodes are $\{-\sqrt{3/2}, 0, \sqrt{3/2}\}$.
The criterium to derive the weights is rather obvious: the moments of the discrete variable have to be equal to the moments of the original continuous variable. The number of matched moments depends on the number of nodes and the moment matching condition is given by

$$\int_a^b H_k(x)\,p\,(x)\,dx = \sum_{i=1}^{n} \omega_i\,H_k(x_i) \qquad \text{for } k = 0, \ldots, n-1. \tag{4.16}$$

For $k = 0$ the Hermite polynomial is 1 and the first condition requires the weights to sum to 1 as an obvious need for a univariate density approximation. For $k = 1$ the Hermite polynomial is $2\,x$ and the condition matches the first raw moment. For $k = 2$ the second is matched and so on. Another interesting interpretation of the moment matching condition is related to the polynomial approximation viewpoint of integration. In equation (4.16) we can see that the expected value of a polynomial transformation will be exact. The number of nodes determines for which maximal degree of polynomial transformation the approximation of the expected value will be exact. Or put it differently, if a high degree polynomial is

needed for an approximation of the integrand then its integration will need many nodes. The moment matching condition can be further simplified to

$$\begin{aligned}\int_a^b p(x)\,dx = 1 &= \omega_1 + \cdots + \omega_n && \text{for } k=0 \text{ by } H_0(x)=1\\ 0 &= \omega_1 H_k(x_1) + \cdots + \omega_n H_k(x_n) && \text{for } k>0 \text{ by } \langle H_k, H_0\rangle = 0\end{aligned}$$

and summarized in a linear system for the weights $\boldsymbol{\omega}$

$$\begin{aligned}\boldsymbol{\Psi} &= \boldsymbol{\Phi}\,\boldsymbol{\omega}\\ \begin{pmatrix}1\\0\\\vdots\end{pmatrix} &= \begin{pmatrix}H_0(x_1) & \cdots & H_0(x_n)\\ \vdots & \ddots & \vdots\\ H_{n-1}(x_1) & \cdots & H_{n-1}(x_n)\end{pmatrix}\begin{pmatrix}\omega_1\\\vdots\\\omega_n\end{pmatrix}.\end{aligned}$$

The problem with Gaussian quadrature is that it is defined for univariate integrals and it is silent about the extension of the operator to the multivariate case. The obvious extension is to match the moments of the multivariate random variable and the question is how to combine nodes in one dimension to obtain nodes in many dimensions. The simplest approach is to discretize each of the involved random variables separately and combine the univariate nodes and weights by a Kronecker product rule. This and the Smolyak extension are described in section 4.3.5.

The rational expectations in the general model are formed over the integrand $\boldsymbol{h}$. $p(x)$ is then a multivariate density of a vector of random shocks $\boldsymbol{e}'$ and the integral is

$$\begin{aligned}\boldsymbol{z} &= \int \boldsymbol{h}(\boldsymbol{s},\boldsymbol{x},\quad \boldsymbol{e}',\boldsymbol{s}',\qquad\qquad \boldsymbol{x}'\qquad\qquad))\,p(\boldsymbol{e}')\,d\boldsymbol{e}'\\ &= \int \boldsymbol{h}(\boldsymbol{s},\boldsymbol{x}^*(\boldsymbol{s}),\boldsymbol{e}',\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x}^*(\boldsymbol{s}),\boldsymbol{e}'),\boldsymbol{x}^*(\boldsymbol{g}(\boldsymbol{s},\boldsymbol{x}^*(\boldsymbol{s}),\boldsymbol{e}'))\,p(\boldsymbol{e}')\,d\boldsymbol{e}'.\end{aligned}$$

4.3.5 Operators

Usually the functions we are approximating or integrating depend on several variables. Therefore we need a rule to extend univariate operators to many dimensions. The usual and most simple approach is to do it by the Kronecker operator. It combines each of the univariate nodes and basis functions with all the others.

An intuition about how to deal with that exponentially growing computational burden more carefully is to take a look at a Taylor series expansion of a function. A second order expansion of a bivariate function around (x^*, y^*) is given by

$$\begin{aligned} F(x,y) &\approx F(x^*,y^*) \\ &+ F_x(x^*,y^*)(x-x^*) + F_y(x^*,y^*)(y-y^*) \\ &+ \frac{1}{2}[F_{xx}(x^*,y^*)(x-x^*)^2 + 2F_{xy}(x^*,y^*)(x-x^*)(y-y^*) \\ &+ F_{yy}(x^*,y^*)(y-y^*)^2] \end{aligned}$$

It can be further simplified by combining the constant terms x^*, y^* and $F_{..}(x^*, y^*)$

$$F(x,y) \approx \alpha_0 + \alpha_1 x + \alpha_2 y + \alpha_3 xy + \alpha_4 x^2 + \alpha_5 y^2.$$

This second order bivariate Taylor expansion is not derived by the Kronecker product of second order univariate polynomials. This would be a linear combination of the polynomials

$$\{1, x, y, xy, x^2, y^2, x^2y, xy^2, x^2y^2\} = \{1, x, x^2\} \otimes \{1, y, y^2\}.$$

The important message is that we can drop the elements x^2y, xy^2, x^2y^2 without loosing asymptotic accuracy. The resulting polynomials are called complete polynomials and are characterized by the property that terms only up to a certain sum of the powers of the involved univariate polynomials are used. In the example we need only basis functions of the Kronecker product $x^a y^b$ where $0 \leq a + b \leq 2$.
An important questions for approximation and integration is how to combine nodes in one dimensions to form nodes in many dimensions. For Kronecker operations on functions the associated operation on the one dimensional nodes is the simple Cartesian product. The optimal multidimensional nodes for complete polynomials are not obvious without the Smolyak operator. Since the operator gives complete polynomials as a special case it can also be used to derive the associated optimal grid. The principle of using only some combinations of Kronecker products was formulated as early as Smolyak (1963a). It was recently used by Krüger and Kübler (2004) to solve an OLG model.
The operator also extends finite element approximation of any degree to many dimensions. Since function approximation and integration are fundamentally the same operators, the Smolyak operator can be applied to both operators without being subject to the curse of dimensionality.
Another very useful property of the Smolyak algorithm is that it constructs the approximation hierarchically. This feature gives an approximation accuracy estimate as a by-product and will be discussed in section 4.3.5.3.

4.3.5.1 Kronecker

The univariate ($d = 1$) approximation operator for a function $f : [0, 1] \to \mathbb{R}$ is

$$U^i(f) = \sum_{j=1}^{m_i} a_j^i \, f(x_j^i)$$

where $i \in \mathbb{N}$, $a_j^i \in \mathbb{C}([-1, 1])$ and $x_j^i \in [-1, 1]$ for a normed approximation. In case of approximating an integration operator, a_j^i are the weights and therefore real numbers and in case of function approximation they are functions implicitly containing the polynomial coefficients. The multidimensional ($d > 1$) tensor or Kronecker product operator $\otimes$ is given by

$$(U^{i_1} \otimes \cdots \otimes U^{i_d})(f) = \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_d=1}^{m_{i_d}} (a_{j_1}^{i_1} \otimes \cdots \otimes a_{j_d}^{i_d}) f(x_{j_1}^{i_1}, ..., x_{j_d}^{i_d}).$$

In this algorithm $m_{i_1} \cdots m_{i_d}$ function evaluations are needed to construct the approximation or integration. This establishes the curse of dimensionality of the Kronecker operator.

4.3.5.2 Smolyak

The Smolyak operator is an active topic in numerics and especially complexity research, see for example Gerstner and Griebel (2003a), Novak and Ritter (1999) or Wasilkowski and Woźniakowski (1999). It is a recursive construction given by

$$A_{q,d}(f) = \sum_{|\,\mathbf{i}\,| \le q} (\Delta^{i_1} \otimes \cdots \otimes \Delta^{i_d})(f) = A_{q-1,d}(f) + \underbrace{\sum_{|\,\mathbf{i}\,| = q} (\Delta^{i_1} \otimes \cdots \otimes \Delta^{i_d})(f)}_{\Delta A_{q,d}(f)} \quad (4.17)$$

where $A_{d-1,d} = 0$, $U^0 = 0$, $\Delta^i = U^i - U^{i-1}$. $\mathbf{i}$ is a d-dimensional vector with norm $|\,\mathbf{i}\,| = i_1 + \cdots + i_d$. $q \ge d$ drives the approximation accuracy. This formulation does not repeat polynomial terms in the various U^i and is more suited for programming purposes. The following formulation reveals the underlying structure more clearly

$$A_{q,d}(f) = \sum_{q-d+1 \le |\,\mathbf{i}\,| \le q} (-1)^{q-|\,\mathbf{i}\,|} \binom{d-1}{q-|\,\mathbf{i}\,|} (U^{i_1} \otimes \cdots \otimes U^{i_d}). \quad (4.18)$$

The Smolyak operator is a linear combination of certain low level Kronecker product operators. The recursive structure in equation (4.17) suggests that the approximation can be refined by increasing q. At each level the approximation gain can be calculated. In case of a function approximation the gain is given by the difference between the interpolation and the true function value at the finer grid. In case of an integration the accuracy gain is simply the change of

the integral value. Both values, if small enough, can be used to stop at a certain accuracy before a maximal approximation degree is reached.
Another rule to reduce the curse of dimensionality is the Gaussian quadrature algorithm of Genz and Keister (1996b). It is an application of the fully symmetric method of Genz (1986) to Patterson (1968) extensions of Gaussian quadrature for normally distributed shocks. Novak and Ritter (1999) show that the fully symmetric algorithm is a special case of the Smolyak operator. I use it for the rational expectation and filtering integrals since Patterson (1968) extensions use nested univariate nodes and thus allow for adaptive integration. Any other nonnested quadrature scheme can be used for nonadaptive integration. Heiss and Winschel (2005) demonstrated the operator's potential in microeconometrics where Smolyak Gaussian quadrature of nonlinear likelihood integrals in a mixed logit model dramatically outperformed Monte Carlo methods up to twenty dimensions. The operator establishes Gaussian quadrature as a competitor to simulation methods for high dimensions and smooth functions. Common wisdom was that integrals beyond five dimensions can be solved only by Monte Carlo methods. The basic operator is competitive at least up to 30 dimensions. Further refinement is possible and integrals with a special structure and several hundred dimensions were solved in Gerstner and Griebel (2003a).
The Smolyak rule operates on a set of one dimensional nodes analogously to operations on functions in order to construct the multidimensional nodes. These nodes are the points where f is evaluated in order to construct $A_{q,d}(f)$. The one dimensional nodes $X^{i_1},..,X^{i_d}$, $X^i=\{x_1^i,...,x_{m_i}^i\}$, used by U^i, are not combined by the Cartesian product $\times$. Instead again only some certain combinations are used

$$H_{q,d}=\bigcup_{q-d+1\leq|\mathbf{i}|\leq q}\left(X^{i_1}\times\cdots\times X^{i_d}\right). \tag{4.19}$$

For a recursive approximation refinement, the function evaluations in one level should be reused in the next level. The one dimensional nodes have then to be selected in a nested way so that $X^i\subset X^{i+1}$. This implies $(X^i\times X^j)\subset(X^i\times X^{j+1})$ and therefore

$$H_{q,d}\subset H_{q+1,d}. \tag{4.20}$$

Equation (4.19) can be rewritten recursively

$$H_{q,d} = \bigcup_{|\mathbf{i}|\leq q}\left(X_\Delta^{i_1}\times\cdots\times X_\Delta^{i_d}\right) \tag{4.21}$$

$$\Delta H_{q,d} = \bigcup_{|\mathbf{i}|=q}\left(X_\Delta^{i_1}\times\cdots\times X_\Delta^{i_d}\right) \tag{4.22}$$

$$H_{q,d} = H_{q-1,d}\cup\Delta H_{q,d}. \tag{4.23}$$

with $X^0=\emptyset$, $X_\Delta^i=X^i\setminus X^{i-1}$ and $H_{d-1,d}=\emptyset$.

The grid used in this work is the Gauss-Lobatto grid defined by

$$m_i = \begin{cases} 1 & \text{for } i = 1 \\ 2^{i-1} + 1 & \text{for } i > 1 \end{cases}$$

$$x_j^i = \begin{cases} 0 & \text{for } i = 1 \\ -\cos\left(\frac{\pi(j-1)}{m_i - 1}\right) & \text{for } j = 1, ..., m_i. \end{cases}$$

These nodes are different than the roots of Chebyshev polynomials. They deliver similar optimal results but have the advantage to be nested.
The following examples are meant to clarify the unusual notation and therefore apparently complicated structure of the Smolyak formulas. The key is the summation set of the index vectors **i** in formula (4.18). In case of a two dimensional ($d = 2$) operator these are the following sets of vectors

$$\begin{aligned} q = 2 : \{\mathbf{i} : 2 - 2 + 1 = 1 \leq |\,\mathbf{i}\,| \leq 2\} &= \{[1\ 1]\} \\ q = 3 : \{\mathbf{i} : 3 - 2 + 1 = 2 \leq |\,\mathbf{i}\,| \leq 3\} &= \{[1\ 1], [1\ 2], [2\ 1]\} \\ q = 4 : \{\mathbf{i} : 4 - 2 + 1 = 3 \leq |\,\mathbf{i}\,| \leq 4\} &= \{[1\ 2], [2\ 1], [1\ 3], [3\ 1], [2\ 2]\} \end{aligned}$$

This gives the following two dimensional Chebyshev polynomials in x and y for the levels $q = 2$ and $q = 3$

$$\begin{aligned} q = 2 \quad &: \quad \{(c_1^1) \times (c_1^2)\} \\ q = 3 \quad &: \quad \{(c_1^1) \times (c_1^2)\} \\ &\quad\ \{(c_1^1) \times (c_1^2 + c_2^2 y + c_3^2 (2y^2 - 1))\} \\ &\quad\ \{(c_1^1 + c_2^1 x + c_3^1 (2x^2 - 1)) \times (c_1^2)\}. \end{aligned}$$

The recurring polynomial terms are left out at subsequent levels in the representation in equation (4.17).
The Smolyak operator on the one dimensional nodes

$$\begin{aligned} &i = 1,\ m_i = 1,\ X^i = \{0\} \\ &i = 2,\ m_i = 3,\ X^i = \{-1, 0, 1\} \\ &i = 3,\ m_i = 5,\ X^i = \{-1, -2^{-0.5}, 0, 2^{-0.5}, 1\}. \end{aligned}$$

constructs the first level of a two dimensional grid as

$$\begin{aligned} H_{2,2} &= (X^1 \times X^1) \\ &= [0\ 0]. \end{aligned}$$

Figure 4.3: Gauss-Lobatto grid at Levels k

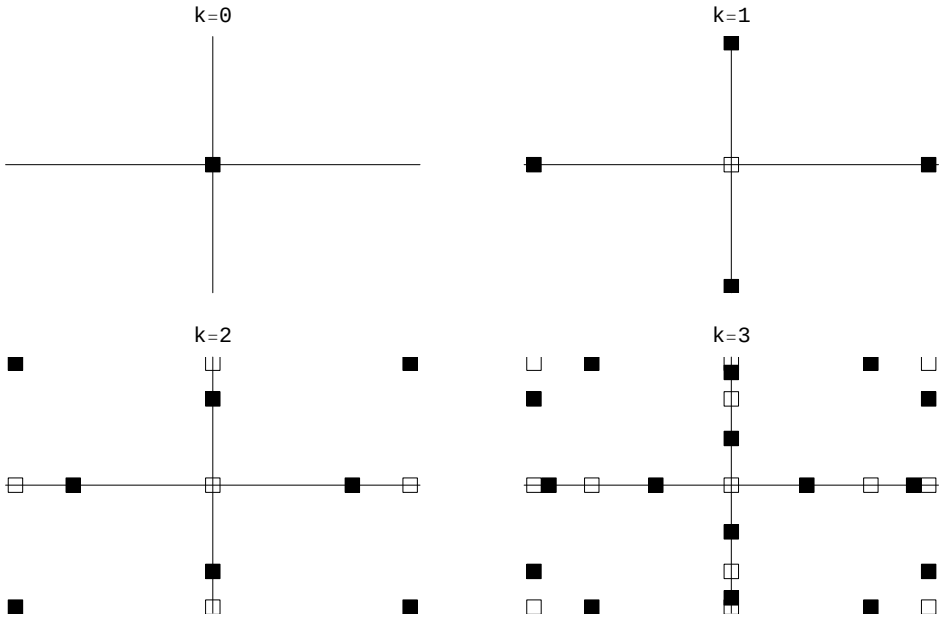
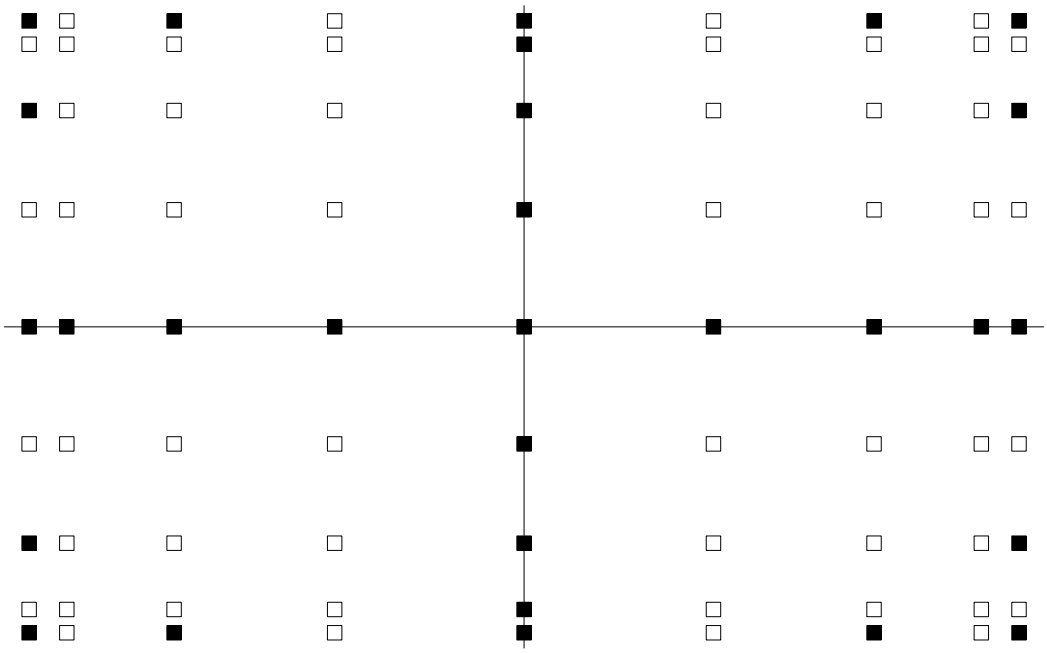


Figure 4.4: Gauss-Lobatto versus Kronecker Product Grid



The grid for the next level is

$$\begin{aligned}
H_{3,2} &\overset{(4.19)}{=} (X^1 \times X^1) \cup (X^1 \times X^2) \cup (X^2 \times X^1) \\
&\overset{(4.20)}{=} (X^1 \times X^2) \cup (X^2 \times X^1) \\
&= \{[0\ -1], [0\ 0], [0\ 1]\} \cup \{[-1\ 0], [0\ 0], [1\ 0]\} \\
&\overset{(4.21)}{=} (X^1_\Delta \times X^1_\Delta) \cup (X^1_\Delta \times X^2_\Delta) \cup (X^2_\Delta \times X^1_\Delta) \\
&\overset{(4.23)}{=} (X^1 \times X^1) \cup (X^1_\Delta \times X^2_\Delta) \cup (X^2_\Delta \times X^1_\Delta) \\
&= \{[0\ 0]\} \cup \{[0\ -1], [0\ 1]\} \cup \{[-1\ 0], [1\ 0]\}
\end{aligned}$$

where the numbers over the equality signs refer to the applied equation. The node [0 0] is contained in the first level grid and appears for nested univariate nodes also in the second level grid. The difference form in equation (4.21) and (4.23) avoids this recurrence and constructs only additional nodes for each level. The grid for the next level is given by

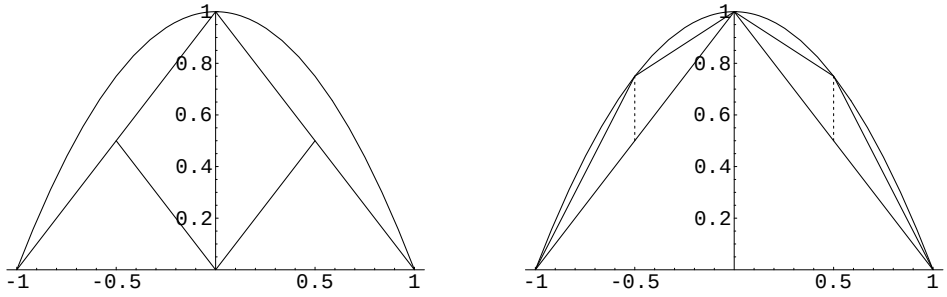
$$H_{4,2} = (X^1 \times X^3) \cup (X^2 \times X^2) \cup (X^3 \times X^1).$$

Figure 4.3 shows the grid for level $k \equiv q - d = 0, 1, 2, 3$ where the filled squares are the new points of the level. Figure 4.4 contrasts the Gauss-Lobatto grid to the tensor product grid.

4.3.5.3 Adaptivity

An interesting feature of the recursive or hierarchical nature of the Smolyak operator is that a convergence criterion is available almost for free. The intuition can be seen most easily in Archimedes' approximation strategy for $f(x) = 1 - x^2$ in figure 4.5. This strategy we would call today hierarchical linear splines.
In the left picture the first and the second level basis functions are plotted together with the function to be approximated. The right hand side shows the linear combination of both levels' basis functions. The two dotted vertical lines are the ap-

Figure 4.5: Hierarchical Surpluses



proximation gains called hierarchical surpluses. They are constructed by approximating the function on the coarse grid $H_{1,1} = \{[-1], [0], [1]\}$ and interpolating the function values at the additional points of the finer grid $\Delta H_{2,1} = \{[-0.5], [0.5]\}$. The difference between these interpolations and the true function values at $\Delta H_{2,1}$ is the surplus. The convergence of the surpluses to zero for ever finer grids makes them a natural convergence criterion. Therefore it is possible to construct algorithms with a stopping criterion once a specified accuracy level is reached. Some structural parameter vectors may imply more or less linear policy functions and an à priori specified level of approximation is inefficient.
This kind of adaptivity approximates up to a certain accuracy for all dimensions. It is already an integral part of the Smolyak operator due to its hierarchical structure. I have implemented it for the integration operator. It is rather simple for nested univariate nodes since the Smolyak algorithm adds in each level some new nodes where the integrand has to be evaluated. The evaluations at lower level nodes can be reused and have only to be reweighted.
There is another more complicated adaptive scheme discussed by Gerstner and Griebel (2003a) which approximates with different accuracy in each dimension. They approximate a 360 dimensional integral in two minutes on a 400 MHz Pentium machine. One of their examples is from economics where a collateral mortgage obligation is priced by a present value integral. The next period realization is discounted only once whereas later realization are discounted more often and contribute less to the present value. It is therefore economical to approximate the next period realization more accurate than the realization in the more distant future. This dimension adaptive approximation scheme is not implemented in the current code.

4.3.6 Weighted Residuals

According to McGrattan (1998) there are three broad approaches to identify the parameter $\boldsymbol{c}$ of the approximated policy function. Common to all of them is that the integral of weighted residuals over the state has to be zero. The methods are characterized by different weight functions.
The residual in the economic application is the value of the first order condition with the approximated policy functions substituted for the policy variable. For the general model the residual is given by

$$\begin{aligned} \boldsymbol{r}^*(\boldsymbol{s}, \boldsymbol{x}^*) &= \boldsymbol{f}(\boldsymbol{s}, \boldsymbol{x}^*(\boldsymbol{s}), \sum_j w_j \boldsymbol{h}(\boldsymbol{s}, \boldsymbol{x}^*(\boldsymbol{s}), \boldsymbol{e}'_j, \boldsymbol{s}'_j, \boldsymbol{x}^*(\boldsymbol{s}'_j)) \\ \boldsymbol{s}'_j &= \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}^*(\boldsymbol{s}), \boldsymbol{e}'_j). \end{aligned}$$

The policy function is defined by its coefficients $\boldsymbol{x}^*(\boldsymbol{s}) \equiv \boldsymbol{x}^*(\boldsymbol{s}; \boldsymbol{c})$ and the residual can alternatively be written as

$$\boldsymbol{r}(\boldsymbol{s}, \boldsymbol{c}) \equiv \boldsymbol{r}^*(\boldsymbol{s}, \boldsymbol{x}^*).$$

This form is more appropriate in order to highlight the implications of the identifying conditions for policy coefficient vector $\boldsymbol{c} = (c_1, ..., c_n)$. The n conditions require the n coefficients to be chosen such that the weighted residuals integrate over the states to zero

$$\int_{\boldsymbol{S}} w_i(\boldsymbol{s})\, \boldsymbol{r}(\boldsymbol{s}, \boldsymbol{c})\, d\boldsymbol{s} = 0 \qquad \text{for} \qquad i = 1, ..., n. \tag{4.24}$$

The Galerkin scheme determines the coefficients so that all available structure in the residuals is used. This requires the residuals to be orthogonal to the basis functions and the weights are therefore the basis functions $w_i(\boldsymbol{s}) = \psi_i(\boldsymbol{s})$ in equation (4.24).
The least squares method requires to minimize

$$\min_{\boldsymbol{c}} \int_{\boldsymbol{S}} \boldsymbol{r}^2(\boldsymbol{s}, \boldsymbol{c})\, d\boldsymbol{s}$$

and the general form (4.24) is a first order condition for minimal squared residuals with weights $w_i(\boldsymbol{s}) = \partial\, \boldsymbol{r}(\boldsymbol{s}, \boldsymbol{c}) / \partial\, c_i$. For both methods multidimensional integrals have to be evaluated. The Galerkin method is the optimal choice for finite elements whereas the collocation method is used for spectral approximation. It specifies the weights as $w_i(\boldsymbol{s}) = \delta(\boldsymbol{s} - \boldsymbol{s}_i)$ where δ is again the Dirac delta function. Substituting the weights in the general form we see that the n identification conditions for $\boldsymbol{c}$ forces the residual to be zero at the nodes

$$\int_{\boldsymbol{S}} \delta(\boldsymbol{s} - \boldsymbol{s}_i)\, \boldsymbol{r}(\boldsymbol{s}, \boldsymbol{c})\, d\boldsymbol{s} = \boldsymbol{r}(\boldsymbol{s}_i, \boldsymbol{c}) = 0 \qquad \text{for} \qquad i = 1, ..., n$$

where $\boldsymbol{s}_i$ is one of n multidimensional nodes.

4.3.7 Implicit Policy Function

So far the discussion of the approximation methods assumed that the function values at the nodes can be obtained by simply evaluating the function we are approximating. This is not the case in functional equations since the function we want to approximate is the one we are looking for. That means we have to approximate functions we do not know but which have to fulfill the first order conditions. The solution to this problem is an iterative procedure with an initial guess and subsequent refinement until the first order condition at the nodes and their associated policy values is zero. The whole policy function then consists of the exact policy values at the nodes and the interpolated policy elsewhere. Solving for the nonlinear rational expectation equilibrium is therefore a nonlinear root finding procedure with an integral evaluation for the rational expectations along the way to evaluate the residuals. In the Galerkin and least squares procedure the root has to be found in the weighted residuals and in case of the collocation approach in the residuals at the nodes.

Table 4.2: Implicit Function Iteration

0. choose initial policy values $\boldsymbol{x}^{(0)}$ at grid $\boldsymbol{s}^G$
1. approximate policy function $\boldsymbol{c}^{(k)} = \boldsymbol{\Psi}\left(\boldsymbol{s}^G\right)^{-1}\boldsymbol{x}^{(k)}$
2. calculate rational expectations
   (a) for all discrete shock realizations $j = 0, ..., J$
      i. $\boldsymbol{s'}_j = \boldsymbol{g}(\boldsymbol{s}^G, \boldsymbol{x}^{(k)}, \boldsymbol{e'}_j)$
      ii. $\boldsymbol{x'}_j = \boldsymbol{\Psi}(\boldsymbol{s'}_j)\,\boldsymbol{c}^{(k)}$
   (b) weight future realizations $\boldsymbol{z} = \sum_j w_j \boldsymbol{h}(\boldsymbol{s}^G, \boldsymbol{x}, \boldsymbol{e'}_j, \boldsymbol{s'}_j, \boldsymbol{x'}_j)$
3. solve for expectational policy $\boldsymbol{x}^e$ in $\boldsymbol{f}(\boldsymbol{s}^G, \boldsymbol{x}^e, \boldsymbol{z}) = 0$
4. update policy by
   (a) function iteration $\boldsymbol{x}^{(k+1)} = \alpha\,\boldsymbol{x}^{(k)} + (1-\alpha)\,\boldsymbol{x}^{\boldsymbol{e}}$ or
   (b) root finding $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - [\partial\,\boldsymbol{r}(\boldsymbol{s}^G, \boldsymbol{x}^{(k)})/\partial\,\boldsymbol{x}]^{-1}\boldsymbol{r}(\boldsymbol{s}^G, \boldsymbol{x}^{(k)})$
5. $k = k + 1$
6. do 1-6 while $\boldsymbol{r}(\boldsymbol{s}^G, \boldsymbol{x}^{(k)}) \neq 0$

I have implemented the collocation approach with two iterative schemes. The first is a fast Newton type root finding algorithm and the second is a robust function iteration. The root finding scheme is a standard Newton iteration given by

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \left(\frac{\partial\,\boldsymbol{r}(\boldsymbol{s}, \boldsymbol{x}^{(k)})}{\partial\,\boldsymbol{x}}\right)^{-1}\boldsymbol{r}(\boldsymbol{s}, \boldsymbol{x}^{(k)})$$

where $\boldsymbol{x}^{(k)}$ are the policy function values at the nodes in iteration k. Since the policy function can be either represented by the policy values $\boldsymbol{x}^{(k)}$ or the implied function coefficients $\boldsymbol{c}^{(k)}$ the function iteration can also be done over the coefficient vector. I iterate over the policy values since some coefficients can be close to zero and the algorithm may become unstable.
The steps for the calculation of the approximation parameters or the policy function at the state grid $\boldsymbol{s}^G$ are summarized in box 4.2. The first step is to start with some policy values at the state grid in step 0. Step 1 calculates the policy coefficients for the policy at the grid. They are needed for the interpolation in step 2 (a) ii where the next period policy at the state realization has to be cal-

culated. Step 2 evaluates the rational expectations. This is done by calculating for each of the possible discrete shocks $\boldsymbol{e}_j$ the next period state in 2(a)i and the next period policy in 2(a)ii. With next period states and policies for all possible discrete shocks we can evaluate the rational expectations $\mathrm{E}\,\boldsymbol{h}$ by weighting the possible future paths in step 2(b). This gives the expectational variables $\boldsymbol{z}$. In step 3 the expectational policy $\boldsymbol{x}^e$ is derived for a given state grid and rational expectation variable $\boldsymbol{z}$ by an inner root finding algorithm. For some models, like the example model, $\boldsymbol{f}(\boldsymbol{s}^G, \boldsymbol{x}^e, \boldsymbol{z}) = 0$ can be solved explicitly for $\boldsymbol{x}^e$ and the inner root finding is not necessary. The optimal policy is found if the expectational policy is the same $\boldsymbol{x}^e = \boldsymbol{x}^{(k)}$ as the one used to generate the expectational variables $\boldsymbol{z}(\boldsymbol{x}^{(k)})$. This is another way to say that the residual for this policy has to be zero $\boldsymbol{f}(\boldsymbol{s}, \boldsymbol{x}^{(k)}, \boldsymbol{z}(\boldsymbol{x}^{(k)})) = \boldsymbol{0}$. If this is not the case the policy for the next iteration is generated in step 4. The function iteration algorithm in 4(a) determines it as a mixture of the last and the expectational solution. In the example model damping was not necessary and I took $\alpha = 0$ for the fastest convergence so that the expectational solution is taken as the policy for the next iteration. In case of the Newton algorithm in 4(b) the next policy is calculated according to the Jacobian of the residuals. Both methods differ only in step 4 where a policy for the next iteration is calculated. They repeat steps 1.-6. until a sufficiently accurate solution with (weighted) residuals close to zero is found.
The iteration is described for a Kronecker product approximation in the approximation step 1 and the interpolation step 2(a)ii. The Smolyak operator is a linear combination of the Kronecker operator and the procedure is therefore analogous. A finite elements approximation strategy needs a different step 2(a)ii where the interpolation of the next period policy is calculated. Moreover the state grid is different since for finite elements the grid can be chosen freely. If the Galerkin weighted residuals are used it changes step 1 where the approximation parameters are determined.
For a large grid the Jacobian is costly to calculate and I use Broyden's variant of the Newton algorithm. The inverse of the Jacobian $[\partial\, r(\boldsymbol{s}, \boldsymbol{x}^{(k)})/\partial\, \boldsymbol{x}]^{-1}$ is approximated by a matrix $\boldsymbol{A}^{(k)}$ and updated according to the residual starting with an initial identity matrix.
The trade off between the function iteration and the Broyden algorithm is that the Jacobian based steps converge faster but the function iteration is more robust for any start values. I implemented both methods. The function iteration is used while searching for the modes of the posterior density with the solution of the linearized model as starting values for the policies. In subsequent likelihood evaluations when sampling around the posterior mode the structural parameters and policies do not change much. Then it is save to use Broyden's method with the nonlinear policy at the previous structural parameters as starting values. This implies that calculating the solution takes only a small fraction of the complete running time of the estimation process. The most time consuming part is the

evaluation of the likelihood for a given nonlinear policy. This may change for larger models.

4.3.8 Efficient Calculation

In the estimation procedure some speed reduction ideas can be exploited. For repeated likelihood evaluations the model has to be solved very often. Some parts of the integration and approximation algorithms can be factored out and done only once.
The first saving is rather simple and obvious and allows to compute Gauss-Hermite nodes and weights only once. In each iteration over the structural parameters within the estimation process, the covariance matrix of the involved shocks changes and implies different nodes for the approximation of the rational expectation integral. This is also true for the evaluation of the likelihood where moments of nonlinear transformations of normal random variables with changing moments have to be calculated. Since the Gauss-Hermite formulas deliver nodes and weights for a standard normal density function, changing moments can be taken into account by a simple change of variables. Nodes $\boldsymbol{y}^{(i)}$ for the standard normal density can be transformed by $\boldsymbol{x}^{(i)} = \boldsymbol{\mu} + \boldsymbol{y}^{(i)}\boldsymbol{\Sigma}^{1/2}$ to obtain nodes for the general density $\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ of $\boldsymbol{x}$, where $\boldsymbol{\Sigma}^{1/2}$ is the Cholesky factor of $\boldsymbol{\Sigma}$.
Possible acceleration schemes for the policy function approximation are more involved. Before I present a method for efficiency gains in the matrix variant of approximation in equation (4.15), I want to discuss a method which takes advantage of the known closed form expressions for the Chebyshev polynomial coefficients.
A univariate ($d = 1$) approximation is defined by

$$f(x) \approx U^i(f) = \sum_{j=1}^{m_i} c_j^i \, T_{j-1}(x)$$

where m_i is the Clenshaw-Curtis formula $m_i = 2^{i-1} + 1, m_1 = 1$ and $T_{j-1}(x)$ is a Chebyshev polynomial of degree $j-1$. In case of Chebyshev roots as nodes $x_{i,k} = \cos\left(\frac{\pi(k-0.5)}{m_i}\right)$ the closed form expression for the coefficients is

$$c_j^i = \frac{\sum_{k=1}^{m_i} f(x_{i,k}) T_{j-1}(x_{i,k})}{\sum_{k=1}^{m_i} (T_{j-1}(x_{i,k}))^2}$$

and in case of nested Gauss-Lobatto nodes $x_{i,k} = -\cos(\frac{\pi(k-1)}{m_i - 1})$ we have

$$c_j^i = \frac{d_{i,j}}{\max(1/2, m_i - 1)} \sum_{k=1}^{m_i} \frac{1}{3 - d_{i,k}} f(x_{i,k}) T_{j-1}(x_{i,k})$$

where $d_{i,j} = 1$ for $j = 1, m_i$ and $d_{i,j} = 2$ else. Chebyshev polynomials have a trigonometric representation $T_j(x) = \cos(j \cos^{-1} x)$ and a fast way to calculate

the coefficients is the discrete fast Fourier transform. For the usually real valued functions in economics, further simplification applies the real form of the transform, called cosine transform, as described for example in Press, Flannery, Teukolsky, and Vetterling (1988).
The multivariate ($d > 1$) extension with $\mathbf{x} = [x_1 \ \dots \ x_d]$ can be written as

$$f(\mathbf{x}) \approx (V^{i_1} \otimes \cdots \otimes V^{i_d})(f) = \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_d=1}^{m_{i_d}} c_{j_1,\dots,j_d}^{i_1,\dots,i_d} T_{j_1-1}(x_1) \cdots T_{j_d-1}(x_d).$$

In case of Chebyshev roots as nodes the coefficients can be calculated by

$$c_{j_1,\dots,j_d}^{i_1,\dots,i_d} = \frac{\sum_{k_1=1}^{m_{i_1}} \cdots \sum_{k_d=1}^{m_{i_d}} f(\{x_{i_1,k_1}, \dots, x_{i_d,k_d}\}) T_{j_1-1}(x_{i_1,k_1}) \cdots T_{j_d-1}(x_{i_d,k_d})}{(\sum_{k_1=1}^{m_{i_1}} (T_{j_1-1}(x_{i_1,k_1}))^2) \cdots (\sum_{k_d=1}^{m_{i_d}} (T_{j_d-1}(x_{i_d,k_d}))^2)}$$

whereas the Gauss-Lobatto grid results in

$$\begin{aligned} c_{j_1,\dots,j_d}^{i_1,\dots,i_d} &= \frac{d_{i_1,j_1} \cdots d_{i_d,j_d}}{\max(1/2, m_{i_1} - 1) \cdots \max(1/2, m_{i_d} - 1)} \\ &\quad \sum_{k_1=1}^{m_{i_1}} \cdots \sum_{k_d=1}^{m_{i_d}} \frac{f(\{x_{i_1,k_1}, \dots, x_{i_d,k_d}\}) T_{j_1-1}(x_{i_1,k_1}) \cdots T_{j_d-1}(x_{i_d,k_d})}{(3 - d_{i_1,k_1}) \cdots (3 - d_{i_d,k_d})}. \end{aligned}$$

The coefficients of a multivariate approximation can be calculated by repeated application of the univariate discrete Fourier transform. Specialized multivariate transform algorithms drop the calculation costs further to a fraction of around $2/d$.
It is clear that the closed form expressions and the fast Fourier transform replicate the matrix inversion $\mathbf{c} = \mathbf{T}^{-1}(\boldsymbol{s}) f(\boldsymbol{s})$ in equation (4.15). The insight I used is that the inverted Chebyshev matrix $\mathbf{T}^{-1}(\boldsymbol{s})$ in step 1 in table 4.2 is the same for each iteration when solving for the policy function. It is even the same for all likelihood evaluations if one restricts the space of approximation to be the same for all iterations over structural parameters. Therefore the closed form inversion can be outperformed by a large factor if the Chebyshev basis matrices are inverted only once. The calculation of coefficients in each root finding iteration reduces thereby to a simple matrix multiplication $\mathbf{c} = \mathbf{T}^{-1}(\boldsymbol{s}) f(\boldsymbol{s})$. This is even faster than the repeated fast Fourier transform equivalent.
A simplification of the matrix inversion can be achieved if one uses the Kronecker inversion theorem $\boldsymbol{T}^{-1} = (\boldsymbol{T}_1 \otimes \cdots \otimes \boldsymbol{T}_d)^{-1} = \boldsymbol{T}_d^{-1} \otimes \cdots \otimes \boldsymbol{T}_1^{-1}$. This inversion appears in the approximation process for the Kronecker product extension (and therefore in the Smolyak operator) since the multivariate basis matrix $\boldsymbol{T}$ is the Kronecker product of the univariate basis matrices $\boldsymbol{T}_i$. But this cost reduction is of minor importance since the inversion has to be done only once.
The reduction scheme in the matrix variant does not take into account the special structure of the Smolyak operator. It reduces the costs of a product rule

approximation and carries over to the Smolyak operator where product rules are combined linearly. The coefficients $c_{j_1,\ldots,j_d}^{i_1,\ldots,i_d}$ are in fact products of the underlying univariate coefficients. Together with the hierarchical structure of the Smolyak operator and its linear Kronecker product combinations there are many repeated calculations. These efficiency gains are exploited by Petras (1999) who carefully remembers which underlying basic coefficients were already calculated when processing through the Smolyak levels. Again this is a matrix inversion equivalent and it is not clear whether it is faster compared to the simple multiplication with a once and for all inverted basis matrix.
Even after all tricks to lower running times one computer may be too slow for an interesting model. This can be expected to happen in case of heterogeneous agents models like OLG models or international macroeconomic models when several countries with possibly several sectors within each country have to be modelled. A recent development in the computer industry is to connect several standard desktop computers to form a cluster and divide the problem into subproblems which can then be solved by the connected computers in parallel processes. Afterwards the partial results are combined to the overall result. A prerequisite for such a procedure is that the problem can be divided in such a way that the communication needed between the involved computer does not exceed the time advantage of dividing the problem. The methods used should therefore be suitable for parallelization. This is the case for the finite element approximation since the approximation space is effectively divided into subspaces and can therefore be distributed to single computers. For spectral approximation dividing the solution process is more difficult if not impossible. However, the proposed Metropolis-Hastings algorithm allows to parallelize the estimation process even if the approximation cannot be distributed.

4.3.9 Euler Error

The approximation procedure should be accompanied by error estimates to assure convergence and to control the approximation quality. The exact policy functions would give zero first order condition residuals for all states. Each deviation is therefore due to the approximation once the policy function is substituted for the policy variable.
What says a certain number of the residual about the approximation quality? Judd (1992) proposed to normalize the residual which can then be interpreted in an economically meaningful way. This can be done by dividing the residuals of the Euler equation by the marginal utility. For the interpretation it is helpful to do an intermediate step in the model at hand and isolate consumption. Dividing

the residual by $(1-x_l(\boldsymbol{s_t}))^{(1-\theta)(1-\tau)}$ and taking it to the power of $1/(\theta(1-\tau)-1)$ gives the Euler equation in terms of consumption

$$\begin{aligned} r^c(\boldsymbol{s}) &= x^c(\boldsymbol{s}) - \left(\frac{\beta \mathrm{E}_t h(\boldsymbol{s}, \boldsymbol{x}^*(\boldsymbol{s}), \boldsymbol{e}, \boldsymbol{s'}, \boldsymbol{x}^*(\boldsymbol{s'}))}{(1-x^l(\boldsymbol{s}))^{(1-\theta)(1-\tau)}}\right)^{\frac{1}{\theta(1-\tau)-1}} \\ \boldsymbol{s}' &= \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}^*(\boldsymbol{s}), \boldsymbol{e}'). \end{aligned}$$

Dividing this expression by consumption and taking its absolute value

$$r^E = \left|\frac{r^c(\boldsymbol{s})}{\psi_c(\boldsymbol{s})}\right|$$

we obtain the Euler error for a given state vector and a nice interpretation. The Euler error r^E is the fraction of consumption expenditures lost by relying on an approximated rather then exact policy. The logarithm of the error $\log_{10} r^E$ delivers a more informative plot. A log Euler error of -3 says that one has to consume 1000 units before one is lost. The maximal error in the whole approximation space can finally be taken to characterize the approximation quality.

4.4 Likelihood

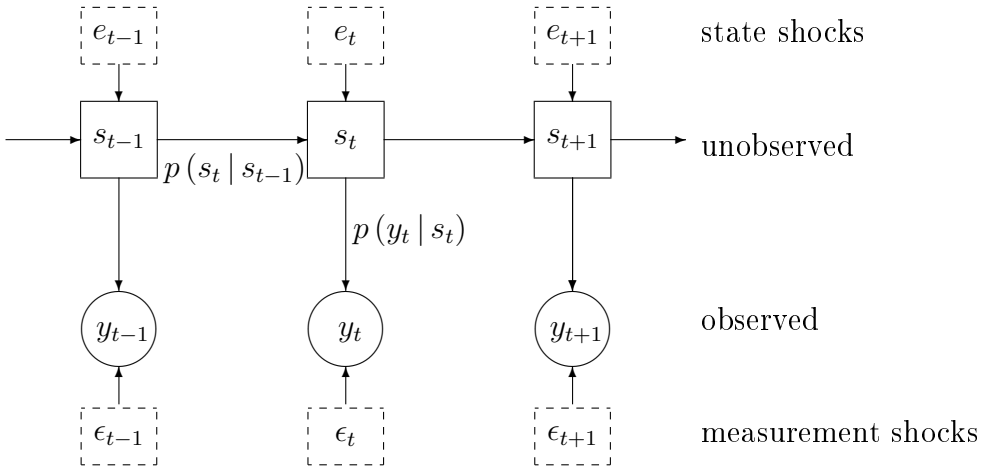
Once the solution of the model $\boldsymbol{x}_t = \boldsymbol{x}^*(\boldsymbol{s}_t)$ is obtained, the implied dynamics can be compared to data. The main problem on the way to analyze the density of observables of the state space model is to derive the density of unobservables. The observables density evaluated at the realizations $\boldsymbol{y}^0$ is functionally equivalent to the likelihood. For a linear state space model with Gaussian shocks the Kalman (1960) filter provides a closed form solution for both densities. Nonlinearity and non-Gaussian distributed state and measurement shocks make it in general impossible to arrive at analytical expressions and approximations are needed.
The state space model is described by the state and measurement equations

$$\boldsymbol{s}_{t+1} = \boldsymbol{g}(\boldsymbol{s}_t, \boldsymbol{x}^*(\boldsymbol{s}_t), \boldsymbol{e}_{t+1}) = \boldsymbol{g}^*(\boldsymbol{s}_t, \boldsymbol{e}_{t+1}) \tag{4.25}$$

$$\boldsymbol{y}_t = \boldsymbol{m}(\boldsymbol{s}_t, \boldsymbol{\epsilon}_t) \tag{4.26}$$

with unobservables $\boldsymbol{s}_t$ and observables $\boldsymbol{y}_t$. State $\boldsymbol{e}_t$ and measurement shocks $\boldsymbol{\epsilon}_t$ follow some distribution. The structure of the state space model is summarized in the Bayes net in figure 4.6. If the policy function is plugged into the state equation it implies the equilibrium transition $\boldsymbol{g}^*$. In the following I will drop the starred notation and the state transition equation with two arguments is implicitly assumed to be the equilibrium transition. In general the model equations can be time dependent, for example through varying variances or other changing parameters.
The autoregressive formulation in equations (4.25) and (4.26) is convenient for the numerical and economic analysis. It describes a nonlinear transformation

Figure 4.6: State Space Model



of the involved random variables and results in some nonstandard densities for the observables and unobservables. For the statistical discussion it is useful to alternatively describe the model by three densities

$$p(\boldsymbol{y}_t \mid \boldsymbol{s}_t) \qquad p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1}) \qquad \text{for } t = 1, ..., T \qquad p(\boldsymbol{s}_0).$$

The first density represents the measurement equation and relates the unobserved to the observed variables. The second density describes the equilibrium state transition in time. The third is the initial information about the state before data is observed. The unobserved state process is Markovian with $p(\boldsymbol{s}_t \mid \boldsymbol{s}_{1:t-1}) = p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1})$ and observations are conditional independent given state $p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1}, \boldsymbol{y}_{1:t-1}) = p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1})$. The notation $\boldsymbol{s}_{1:t}$ is used for a sequence of vectors $\{\boldsymbol{s}_1, ..., \boldsymbol{s}_t\}$. All model related densities are without being noted conditioned on model M_i and the associated structural parameters $\boldsymbol{\theta}_{M_i}$. The full notation will be needed for the model selection criterion in section 4.6.
The state space model is much more general than the various ARMA models. It is built upon the distinction between observed and unobserved variables. This distinction characterizes most economic theories and many econometric complications and offers the advantage of a unified framework. Stochastic trends can be easily modelled so that nonstationary data can be decomposed in an economically sensible way. It offers the statistical model for a unified growth and business cycle literature. Moreover it allows for cointegration relations, missing observations, measurement errors, learning processes or time varying coefficients as in regime switching models as well as time varying shock distributions as in GARCH models.

In engineering applications the density of the unobserved states is often the end of an investigation. The parameters and the functional form of the model are known and represent for example physical laws. The evaluation of these laws in the light of data is not the purpose of the analysis. In the original application of the state space model the unobserved state was the position of a satellite inferred from noisy measurement of a known transition law in the orbit. The econometric applications of unobserved state estimation are for example the Hodrick-Prescott decomposition of the output in a trend and business cycle component. They are not observable separately but have to be inferred from one single output time series. In its usual formulation the HP filter gives an estimate of the unobserved trend as the solution to the minimization problem

$$\min_{y_t^*} \sum_{t=1}^{T} \frac{1}{\sigma_0^2}(y_t - y_t^*)^2 + \frac{1}{\sigma_1^2}(\Delta^2 y_t^*)^2$$

where y is the observed output, y^* is the unobserved trend, σ_0^2 is the variance of the business cycle $y - y^*$ and σ_1^2 is the variance of the growth rate of the trend. The minimization is invariant to a monotone transformation and what matters is $\lambda = \sigma_0^2/\sigma_1^2$. This parameter depends on the frequency of the data and is usually chosen to be 100 for annual, 400 for semi-annual and 1600 for quarterly data. Harvey (1985) shows the state space representation of the HP filter. The measurement equation is the sum of the unobserved trend and business cycle fluctuation

$$y_t = y_t^* + \epsilon_t$$

with $\epsilon_t \sim \mathcal{N}(0, \sigma_0^2)$. The state equations define the growth rate of the trend

$$\begin{aligned} y_t^* &= g_{t-1} + y_{t-1}^* \\ g_t &= g_{t-1} + e_t \end{aligned}$$

with $e_t \sim \mathcal{N}(0, \sigma_0^2/\lambda)$. The advantage of this representation is that it uncovers the explicit assumption for the trend process. In the HP filter the change of the trend follows a random walk. The Kalman filter provides the unobserved trend $y_{1:T}^*$ as the smoothed estimate of the unobservable state. It can then be used to recover the business cycle as the residual in the measurement equation.
Most often the economic focus is on parameter estimates. In the engineering literature the parameter estimation is called joint (state and parameter) estimation or simply the static problem. The static classification is due to the fact that parameters and states are both treated as unobservable random variables and are therefore statistically not fundamentally different. They differ in the fact that parameters are fixed or static whereas unobserved states follow a dynamic process. The most simple approach to parameter estimation is to augment the state vector by the vector of parameters and to transform the parameter estimation into a

problem of the state estimation. It is not clear whether this works analogously for economics where we are interested in the structural and not the reduced form parameters. Doucet, de Freitas, and Gordon (2001) offer an extensive overview of nonlinear filters.
The problems of likelihood evaluation, prediction or unobserved state estimation have to be solved by deriving the density of the unobserved states. I use the filtering algorithms to derive the likelihood of the sample for a given parameter vector.

4.4.1 State Space

The state space form of the economic model is obtained once the policy functions for labor and consumption $x^l(k,a)$ and $x^c(k,a)$ are calculated. In the example model the state equations describe the capital and productivity dynamics

$$\begin{aligned} k_t &= e^{a_{t-1}} k_{t-1}^{\alpha} x^l(k_{t-1}, a_{t-1})^{1-\alpha} - x^c(k_{t-1}, a_{t-1}) + (1-\delta)k_{t-1} \\ a_t &= \rho a_{t-1} + e_t. \end{aligned}$$

The measurement equations for output y_t, labor l_t and investment i_t are

$$\begin{aligned} y_t &= e^{a_t} k_t^{\alpha} x^l(k_t, a_t)^{1-\alpha} + \epsilon_t^y \\ l_t &= x^l(k_t, a_t) + \epsilon_t^l \\ i_t &= e^{a_t} k_t^{\alpha} x^l(k_t, a_t)^{1-\alpha} - x^c(k_t, a_t) + \epsilon_t^i. \end{aligned}$$

These measurement equations do not have the general but an additive noise form with $\boldsymbol{m}(\boldsymbol{s}_t, \boldsymbol{\epsilon}_t) = \boldsymbol{m'}(\boldsymbol{s}_t) + \boldsymbol{\epsilon}_t$. This feature simplifies the derivation of the period contributions to the likelihood. In this model formulation it is necessary to add measurement errors because there is only one driving structural shock but three observables. Without these additional shocks there would be a deterministic functional dependency between the observed variables. An economically more sensible way to solve this problem is to add structural state shocks or to use less observables. Since the economic analysis is not of primary interest in this work I simply add the measurement errors as in Fernádez-Villaverde and Rubio-Ramírez (2004b). Ruge-Murcia (2003) discusses this topic more extensively.

4.4.2 Filtering

The estimation approach is recursive and uses the Markovian conditional independence of the state equation. Each new observation is used to update the information about the unobservables summarized by the posterior density of the unobserved states. For satellite maneuvers for example this recursive structure corresponds to the real situation since noisy measurements of the satellite's position become available during the control process.

The main tool for the filtering problem is again Bayes' formula. Here it describes how the data helps to learn about the unobservables. The prior density of unobservables describes the available information before the new data has been analyzed. The evidence about the observables incorporated in the likelihood informs us by transforming the prior into the posterior. Once the data is processed the posterior becomes the prior with regard to the new data. Therefore the concept of prior and posterior is defined relative to the new data.
The filtering recursion is a two step procedure. The first step is to form a prediction and the second is the filtering step where new information from data is incorporated to modify the prediction. The state vector contains all information about the system and we start with prior information

$$p(\boldsymbol{s}_0) = p(\boldsymbol{s}_0 \mid \boldsymbol{y}_0).$$

The prediction step represents the prior density $p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1})$

$$\begin{aligned} p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1}) &= \int p(\boldsymbol{s}_t, \boldsymbol{s}_{t-1} \mid \boldsymbol{y}_{1:t-1})\, d\boldsymbol{s}_{t-1} && (4.27) \\ &= \int p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1})\, p(\boldsymbol{s}_{t-1} \mid \boldsymbol{y}_{1:t-1})\, d\boldsymbol{s}_{t-1}. && (4.28) \end{aligned}$$

It is formed by the weighted state transition $p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1})$ with the weight given by the last period posterior $p(\boldsymbol{s}_{t-1} \mid \boldsymbol{y}_{1:t-1})$. The next posterior $p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t})$ is obtained in the filtering step and updates the prediction by incorporating new data $\boldsymbol{y}_t$

$$p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t}) = \frac{p(\boldsymbol{s}_t, \boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1})}{p(\boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1})} = l_t^{-1} p(\boldsymbol{y}_t \mid \boldsymbol{s}_t)\, p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1}). \tag{4.29}$$

Therefore these two recursion steps transform one posterior $p(\boldsymbol{s}_{t-1} \mid \boldsymbol{y}_{1:t-1})$ into the next $p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t})$. The filtering equation (4.29) is the result of a repeated application of Bayes' formula.[2] The normalizing constant l_t in the filtering step

$$l_t = \int p(\boldsymbol{y}_t \mid \boldsymbol{s}_t)\, p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1})\, d\boldsymbol{s}_t = p(\boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1}) \tag{4.30}$$

is the period contribution to the likelihood

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}_{1:T}) \equiv p(\boldsymbol{y}_{1:T} \mid \boldsymbol{\theta}) = \prod_{t=1}^{T} p(\boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1}, \boldsymbol{\theta}) = \prod_{t=1}^{T} l_t.$$

---

[2] $p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t}) \overset{(1)}{=} \frac{p(\boldsymbol{y}_{1:t} \mid \boldsymbol{s}_t)\, p(\boldsymbol{s}_t)}{p(\boldsymbol{y}_{1:t})} \overset{(2)}{=} \frac{p(\boldsymbol{y}_t, \boldsymbol{y}_{1:t-1} \mid \boldsymbol{s}_t)\, p(\boldsymbol{s}_t)}{p(\boldsymbol{y}_t, \boldsymbol{y}_{1:t-1})} \overset{(3)}{=} \frac{p(\boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1}, \boldsymbol{s}_t)\, p(\boldsymbol{y}_{1:t-1} \mid \boldsymbol{s}_t)\, p(\boldsymbol{s}_t)}{p(\boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1})\, p(\boldsymbol{y}_{1:t-1}))} \overset{(4)}{=} \frac{p(\boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1}, \boldsymbol{s}_t)\, p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1})\, p(\boldsymbol{y}_{1:t-1})\, p(\boldsymbol{s}_t)}{p(\boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1})\, p(\boldsymbol{y}_{1:t-1})\, p(\boldsymbol{s}_t)} \overset{(5)}{=} \frac{p(\boldsymbol{y}_t \mid \boldsymbol{s}_t)\, p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1})}{p(\boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1})} \overset{(6)}{=} \frac{p(\boldsymbol{y}_t \mid \boldsymbol{s}_t)\, p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1})}{\int p(\boldsymbol{y}_t \mid \boldsymbol{s}_t)\, p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1})\, d\boldsymbol{s}_t}$
where: (1) Bayes' formula, (2) separate: $\boldsymbol{y}_{1:t} = \{\boldsymbol{y}_t, \boldsymbol{y}_{1:t-1}\}$, (3) factorize: $\frac{p(a,b \mid c) = p(a \mid b,c) p(b \mid c)}{p(a,b) = p(a \mid b) p(b)}$, (4) Bayes' formula, (5) cancel terms: $p(\boldsymbol{y}_{1:t-1})\, p(\boldsymbol{s}_t)$, use conditional independence: $p(\boldsymbol{y}_t \mid \boldsymbol{y}_{1:t-1}, \boldsymbol{s}_t) = p(\boldsymbol{y}_t \mid \boldsymbol{s}_t)$

of the complete sample. The state posterior densities $p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t})$ for $t = 1, ..., T$ are either an end in itself as in the HP filter or an instrument to obtain the sample likelihood $\mathcal{L}$. The period contribution to the likelihood in the last equation and in (4.30) is implicitly conditioned on the parameter vector $\boldsymbol{\theta}$. This is without being further noted also true for all model related densities in the rest of this section. The likelihood can be either maximized over the parameter vector or as in this work used to estimate the parameter posterior density $p(\boldsymbol{\theta} \mid \boldsymbol{y}_{1:T})$. This will be discussed in section 4.5.
If we are interested in the estimates of the unobserved states then the posterior density estimates can be further refined. This so called smoothing procedure works recursively backwards in time for $t = T, ..., 1$ where the state densities $p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:T})$ conditioned on the complete sample are derived. Smoothing algorithms are not discussed since the focus is to evaluate the likelihood.
For the likelihood evaluation at least the first two moments of the posterior density of the unobserved states have to be estimated. This involves repeated integration of a density weighted nonlinear function. That means we need to approximate the expected value of a nonlinearly transformed random variable

$$\mathcal{I}(f)_p = E(f(\boldsymbol{s})) = \int f(\boldsymbol{s})\, p(\boldsymbol{s})\, d\boldsymbol{s}. \tag{4.31}$$

The methods used for these integrals are either a Monte Carlo simulation or Gaussian quadrature as discussed in section 4.3.4. In the filtering literature another so called unscented transform is often used. It is a numerical integration technique developed for nonlinear filtering by Julier and Uhlmann (1997), extended in Julier and Uhlmann (2002) and summarized in Julier and Uhlmann (2004). It is a deterministic integration strategy like Gaussian quadrature.
The unscented filter tries to cope with the shortcomings of the traditional extended Kalman filter for nonlinear state space models. The extended Kalman filter uses a first order Taylor approximation of the nonlinear state space equations in order to approximate the first two moments of the nonlinear transformation of a Gaussian random variable. Since these transformed moments can be given in a closed form for the linearized equations, the hope is that they are a reasonable approximation of the true moments of the nonlinear transform. This is in general not true.
The unscented transform constructs nodes and weights in order to approximate the mean and the covariance of a nonlinear transformation instead of approximating the state transition function. The nonlinear function is evaluated at the nodes and a weighted sum is the integral approximation. The derivation of these nodes relies on identifying conditions which match the true moments of a nonlinear transform. Of course the general idea is the same as in Gaussian quadrature, namely that a density is easier to approximate than a general function. In fact the unscented transform uses low level Gaussian quadrature nodes. The exponentially rising computational burden of the Kronecker product quadrature formulas

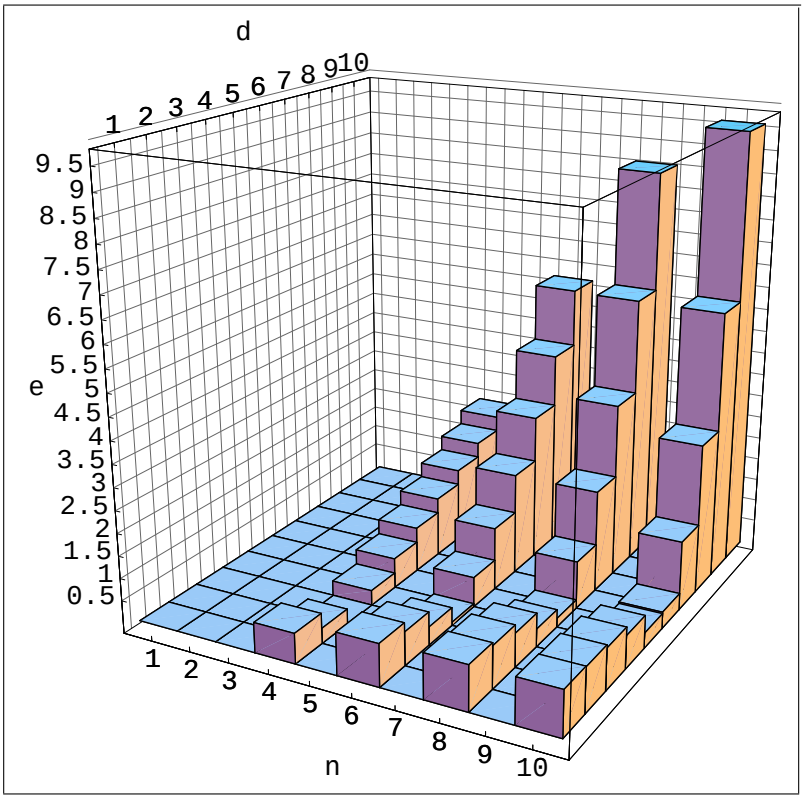
is the usual argument in favor of the unscented transform where the number of nodes is restricted to $2d+1$ and d is the dimension of the random variable.
This restriction of the number of nodes may be a problem in economic models since the policy functions we usually need are of higher polynomial degree. These policy functions are one part of the state equations which have to be integrated in the filtering recursions.
How much decreases the approximation accuracy of the unscented transform with an increasing dimension and polynomial degree of the nonlinear function? If it does not decrease there would be no curse of dimensionality and costs would rise linearly as in $2d+1$. The unscented transform is not a solution of the curse in any respect but merely a shift of the problem from the rising number of nodes for a given accuracy to a given number of nodes with a decreasing accuracy.
In order to gain some intuition for the errors that may occur I calculated an example for a standard normally distributed random variable. I use the unscented nodes and weights to approximate the expected value of a nonlinear transformation of a random variable while increasing the dimension. In order to check the unscented transform's ability to handle different degrees of nonlinearity I take a simple polynomial as integrand. It represents functions with different degree of nonlinearity. The d-dimensional polynomial of degree n is $f_{d,n}(\boldsymbol{x}) = \sum_{i=1}^{d} x_i^n$, where x_i is the i^{th} element of vector $\boldsymbol{x}$. It is integrated over a standard normal uncorrelated random variable with $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \hat{\boldsymbol{x}}, \boldsymbol{\Sigma}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{0}, \boldsymbol{I})$.[3] Since the function is additive and the variables are uncorrelated the true expected value is $\mathcal{I}(f_{d,n})_p = dr^n$ where r^n is the n^{th} raw moment of a standard normally distributed random variable. The first 10 moments are 0, 1, 0, 3, 0, 15, 0, 105, 0, 945. According to Julier and Uhlmann (2004) the unscented nodes for the expected value approximation are given by

$$\begin{aligned} \boldsymbol{x}_j &= \hat{\boldsymbol{x}} + \sqrt{d\boldsymbol{\Sigma}_j} \qquad \text{for } j = 1, ..., d \\ \boldsymbol{x}_j &= \hat{\boldsymbol{x}} - \sqrt{d\boldsymbol{\Sigma}_{j-d}} \qquad \text{for } j = d+1, ..., 2d \end{aligned}$$

with equal weights $w_j = 1/(2d)$. $\boldsymbol{\Sigma}_j$ is the j^{th} column of the Cholesky matrix square root. For a standard normal distribution $\boldsymbol{\Sigma}$ is the identity matrix and the approximation is the simple mean of the function evaluations $\hat{\mathcal{I}}(f_{d,n})_p = \sum_{j=1}^{2d} f(\boldsymbol{x}_j)/(2d)$. The nonlinear function is evaluated at $2d$ vectors with all elements being zero except one which is once $+\sqrt{d}$ and once $-\sqrt{d}$. The approximation of the nonlinear transformation is the mean of the function at these vectors. For the approximation of the covariance the nodes and weights are slightly different. They are parameterized and can capture some non normal higher moments. Since the nonlinear function is symmetric in each variable each evaluation gives $\sqrt{d}^n$. This is also the mean and therefore the approximation is $\hat{\mathcal{I}}(f_{d,n}) = \sqrt{d}^n$ for even n and zero for uneven n. Since the uneven moments of

[3] The notation of a n-variate normal density is $\mathcal{N}(\boldsymbol{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \equiv ((2\pi)^{d_s} \mid \boldsymbol{\Sigma} \mid)^{-1/2} \exp\left(-[\boldsymbol{s}-\boldsymbol{\mu}]^T \boldsymbol{\Sigma}^{-1} [\boldsymbol{s}-\boldsymbol{\mu}]/2\right)$.

Figure 4.7: Relative Error of the Unscented Transform



a normal density are zero and the nodes are symmetric and equally weighted, the approximation will be exact for uneven polynomial degrees. The percentage error of the approximation depends on the dimension and the polynomial degree of the nonlinear function f and is given by $e(d,n) = (\hat{\mathcal{I}} - \mathcal{I})/\mathcal{I}$ for even d and zero for uneven n. Figure 4.7 plots this function. It shows that for a polynomial degree of $n = 2$ the unscented approximation error is zero for any dimension, i.e. the unscented integration has a polynomial exactness of second degree. If a higher polynomial exactness is needed the error rises immediately to .33 and even more for higher dimensions. This approximation strategy will therefore run into problems where polynomial exactness beyond the second degree is needed. The curse of dimensionality is in the unscented transform not an increasing computational burden for a given polynomial exactness but a decreasing accuracy due to an insufficient number of nodes and weights in higher dimensions.
A Smolyak quadrature based filter can be expected to perform better over a wider class of models. It allows a true relieve of the curse of dimensionality with a flexible accuracy according to the needs for a given model. Moreover a Smolyak based Gaussian quadrature filter has the advantage of a more accurate approximation of

non-normal densities whereas in the unscented transform skewness and kurtosis are parameterized in an obscure way to account for non-normal densities.
My focus in the following nonlinear filter discussion is on the question where integration arises and Monte Carlo methods can be substituted by Smolyak based Gaussian quadrature.

4.4.3 Kalman Filter

This section derives the Kalman gain as an exact solution for the linear Gaussian model and as an approximation for nonlinear models. The approximation will be used for the Gaussian (quadrature) filters in later sections.

4.4.3.1 Nonlinear

The nonlinear approximation starts with the decomposition of the posterior density by the Bayes formula

$$p\left(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t}\right) = \frac{p\left(\boldsymbol{s}_t, \boldsymbol{y}_{1:t}\right)}{p\left(\boldsymbol{y}_{1:t}\right)}.$$

For normalized densities we can write $p\left(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t}\right) = p\left(\boldsymbol{s}_t, \boldsymbol{y}_{1:t}\right)$. If we approximate the joint density by the predictive density $p\left(\boldsymbol{s}_t, \boldsymbol{y}_{1:t}\right) \approx p\left(\boldsymbol{s}_t, \boldsymbol{y}_t \,|\, \boldsymbol{s}_{t-1}, \boldsymbol{y}_{1:t-1}\right)$ the Gaussian posterior is given by

$$\begin{aligned} p\left(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t}\right) &= \mathcal{N}(\hat{\boldsymbol{s}}, \boldsymbol{P}^{\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}}) \\ &= \frac{1}{(2\pi)^{d_x/2} \,|\, \boldsymbol{P}^{\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}} \,|^{\,1/2}} \exp(-\frac{1}{2}A) \end{aligned}$$

with

$$\begin{aligned} \hat{\boldsymbol{s}} &= \hat{\boldsymbol{s}}_{t|t} = E(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t}) \\ \boldsymbol{P}^{\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}} &= \boldsymbol{P}^{ss}_{t|t} = E([\boldsymbol{s}_t - \hat{\boldsymbol{s}}][\boldsymbol{s}_t - \hat{\boldsymbol{s}}]^T \,|\, \boldsymbol{y}_{1:t}) \\ A &= [\boldsymbol{s}_t - \hat{\boldsymbol{s}}]^T (\boldsymbol{P}^{\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}})^{-1} [\boldsymbol{s}_t - \hat{\boldsymbol{s}}] \\ &= \boldsymbol{s}_t^T (\boldsymbol{P}^{\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}})^{-1} \boldsymbol{s}_t - \boldsymbol{s}_t^T (\boldsymbol{P}^{\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}})^{-1} \hat{\boldsymbol{s}} \\ &\quad -\hat{\boldsymbol{s}}^T (\boldsymbol{P}^{\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}})^{-1} \boldsymbol{s}_t + \hat{\boldsymbol{s}}^T (\boldsymbol{P}^{\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}})^{-1} \hat{\boldsymbol{s}}. \end{aligned} \tag{4.32}$$

We can define the joint vector

$$\boldsymbol{z}_t \equiv \begin{bmatrix} \boldsymbol{s}_t \\ \boldsymbol{y}_t \end{bmatrix}$$

and assume it to be Gaussian

$$p\left(\boldsymbol{z}_t\right) \sim \mathcal{N}(\bar{\boldsymbol{z}}_t, \boldsymbol{P}^{\boldsymbol{z}\boldsymbol{z}})$$

with

$$\begin{aligned}
\bar{\boldsymbol{z}}_t &= \begin{bmatrix} \bar{\boldsymbol{x}}_t \\ \bar{\boldsymbol{y}}_t \end{bmatrix} = E\left(\begin{bmatrix} \boldsymbol{s}_t \\ \boldsymbol{y}_t \end{bmatrix} \mid \boldsymbol{s}_{t-1}, \boldsymbol{y}_{t-1} \right) = \begin{bmatrix} \hat{\boldsymbol{s}}_{t|t-1} \\ \hat{\boldsymbol{y}}_{t|t-1} \end{bmatrix} \\
\boldsymbol{P}^{\boldsymbol{zz}} &= E\left([\boldsymbol{z}_t - \bar{\boldsymbol{z}}][\boldsymbol{z}_t - \bar{\boldsymbol{z}}]^T \mid \boldsymbol{s}_{t-1}, \boldsymbol{y}_{1:t-1} \right) \\
&= \begin{bmatrix} \boldsymbol{P}^{\boldsymbol{ss}} & \boldsymbol{P}^{\boldsymbol{sy}} \\ \boldsymbol{P}^{\boldsymbol{ys}} & \boldsymbol{P}^{\boldsymbol{yy}} \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1} & \boldsymbol{P}^{\boldsymbol{sy}}_{t|t-1} \\ \boldsymbol{P}^{\boldsymbol{ys}}_{t|t-1} & \boldsymbol{P}^{\boldsymbol{yy}}_{t|t-1} \end{bmatrix}.
\end{aligned}$$

The block inverse of $\boldsymbol{P}^{\boldsymbol{zz}}$ is

$$\begin{aligned}
(\boldsymbol{P}^{\boldsymbol{zz}})^{-1} &= \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{12} & \boldsymbol{B}_{22} \end{bmatrix} \\
&= \begin{bmatrix} (\boldsymbol{P}^{\boldsymbol{ss}} - \boldsymbol{P}^{\boldsymbol{sy}}(\boldsymbol{P}^{\boldsymbol{yy}})^{-1}\boldsymbol{P}^{\boldsymbol{ys}})^{-1} & -\boldsymbol{B}_{11}\boldsymbol{P}^{\boldsymbol{sy}}(\boldsymbol{P}^{\boldsymbol{yy}})^{-1} \\ -\boldsymbol{B}_{22}\boldsymbol{P}^{\boldsymbol{ys}}(\boldsymbol{P}^{\boldsymbol{ss}})^{-1} & (\boldsymbol{P}^{\boldsymbol{yy}} - \boldsymbol{P}^{\boldsymbol{ys}}(\boldsymbol{P}^{\boldsymbol{ss}})^{-1}\boldsymbol{P}^{\boldsymbol{sy}})^{-1} \end{bmatrix}
\end{aligned}$$

and the joint density is given by

$$p\left(\boldsymbol{z}_t\right) = \frac{1}{(2\pi)^{(d_x+d_y)/2} \mid \boldsymbol{P}^{\boldsymbol{zz}} \mid^{1/2}} \exp\left(-\frac{1}{2}C\right)$$

with

$$\begin{aligned}
C &= \begin{bmatrix} (\boldsymbol{s}_t - \bar{\boldsymbol{x}}) & (\boldsymbol{y}_t - \bar{\boldsymbol{y}}) \end{bmatrix}^T (\boldsymbol{P}^{\boldsymbol{zz}})^{-1} \begin{bmatrix} (\boldsymbol{s}_t - \bar{\boldsymbol{x}}) & (\boldsymbol{y}_t - \bar{\boldsymbol{y}}) \end{bmatrix} \\
&= [\boldsymbol{s}_t - \bar{\boldsymbol{x}}]^T B_{11} [\boldsymbol{s}_t - \bar{\boldsymbol{x}}] + [\boldsymbol{s}_t - \bar{\boldsymbol{x}}]^T B_{12} [\boldsymbol{y}_t - \bar{\boldsymbol{y}}] + \\
&\quad [\boldsymbol{y}_t - \bar{\boldsymbol{y}}]^T B_{21} [\boldsymbol{s}_t - \bar{\boldsymbol{x}}] + [(\boldsymbol{y}_t - \bar{\boldsymbol{y}})]^T B_{22} [\boldsymbol{y}_t - \bar{\boldsymbol{y}}] \\
&= \boldsymbol{s}_t^T B_{11} \boldsymbol{s}_t + \boldsymbol{s}_t^T [-B_{11}\bar{\boldsymbol{x}} + B_{12}(\boldsymbol{y}_t - \bar{\boldsymbol{y}})] + ... \qquad (4.33)
\end{aligned}$$

Equating the first terms in equations (4.32) and (4.33) gives

$$\boldsymbol{P}^{\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}} = B_{11}^{-1} = \boldsymbol{P}^{\boldsymbol{ss}} - \boldsymbol{P}^{\boldsymbol{sy}}(\boldsymbol{P}^{\boldsymbol{yy}})^{-1}\boldsymbol{P}^{\boldsymbol{ys}} \qquad (4.34)$$

and from the second terms we obtain

$$\begin{aligned}
(\boldsymbol{P}^{\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}})^{-1}\hat{\boldsymbol{s}} &= B_{11}\bar{\boldsymbol{x}} - B_{12}(\boldsymbol{y}_t - \bar{\boldsymbol{y}}) \\
&= (\boldsymbol{P}^{\boldsymbol{ss}} - \boldsymbol{P}^{\boldsymbol{sy}}(\boldsymbol{P}^{\boldsymbol{yy}})^{-1}\boldsymbol{P}^{\boldsymbol{ys}})^{-1}\bar{\boldsymbol{x}} + \\
&\quad (\boldsymbol{P}^{\boldsymbol{ss}} - \boldsymbol{P}^{\boldsymbol{sy}}(\boldsymbol{P}^{\boldsymbol{yy}})^{-1}\boldsymbol{P}^{\boldsymbol{ys}})^{-1}\boldsymbol{P}^{\boldsymbol{sy}}(\boldsymbol{P}^{\boldsymbol{yy}})^{-1}(\boldsymbol{y}_t - \bar{\boldsymbol{y}})
\end{aligned}$$

which together with equation (4.34) is the solution to the filtering problem

$$\hat{\boldsymbol{s}} = \bar{\boldsymbol{s}} + \boldsymbol{P}^{\boldsymbol{sy}}(\boldsymbol{P}^{\boldsymbol{yy}})^{-1}(\boldsymbol{y}_t - \bar{\boldsymbol{y}}).$$

The Kalman gain

$$\boldsymbol{K}_t = \boldsymbol{P}^{\boldsymbol{sy}}(\boldsymbol{P}^{\boldsymbol{yy}})^{-1}$$

updates recursively the prediction density and provides two parameters characterizing the posterior density

$$\begin{aligned}\hat{\boldsymbol{s}}_{t|t} &= \hat{\boldsymbol{s}}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{y}_t - \bar{\boldsymbol{y}}) \\ \boldsymbol{P}^{\boldsymbol{ss}}_{t|t} &= \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1} - \boldsymbol{K}_t \boldsymbol{P}^{\boldsymbol{yy}}_{t|t-1} \boldsymbol{K}_t^T\end{aligned}$$

with

$$\begin{aligned}\boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1} &= E\left([\boldsymbol{s}_{t|t-1} - \hat{\boldsymbol{s}}_{t|t-1}][\boldsymbol{s}_{t|t-1} - \hat{\boldsymbol{s}}_{t|t-1}]^T\right) \\ \boldsymbol{P}^{\boldsymbol{sy}}_{t|t-1} &= E\left([\boldsymbol{s}_{t|t-1} - \hat{\boldsymbol{s}}_{t|t-1}][\boldsymbol{y}_{t|t-1} - \hat{\boldsymbol{y}}_{t|t-1}]^T\right) \\ \boldsymbol{P}^{\boldsymbol{yy}}_{t|t-1} &= E\left([\boldsymbol{y}_{t|t-1} - \hat{\boldsymbol{y}}_{t|t-1}][\boldsymbol{y}_{t|t-1} - \hat{\boldsymbol{y}}_{t|t-1}]^T\right).\end{aligned}$$

4.4.3.2 Linear

The prediction density for a linear model with Gaussian shocks, a fixed state and measurement covariance matrix $\boldsymbol{Q}$ and $\boldsymbol{R}$ and the prior $p(\boldsymbol{s}_0) = \mathcal{N}(\boldsymbol{s}_0; \hat{\boldsymbol{s}}_0, \boldsymbol{P}^{ss}_0)$ is characterized by its expected value

$$\hat{\boldsymbol{s}}_{t|t-1} = \bar{\boldsymbol{s}} + \boldsymbol{P}(\hat{\boldsymbol{s}}_{\boldsymbol{t-1|t-1}} - \bar{\boldsymbol{s}}).$$

and the covariance

$$\boldsymbol{P}^{ss}_{t|t-1} = \boldsymbol{P}\boldsymbol{P}^{ss}_{t-1|t-1}\boldsymbol{P}' + \boldsymbol{Q}$$

where $\boldsymbol{P}$ is the equilibrium state transition and $\bar{\boldsymbol{s}}$ the steady state. The observable innovation is

$$\boldsymbol{y}^i_t = \boldsymbol{y}_t - (\bar{\boldsymbol{M}} + \boldsymbol{M}\hat{\boldsymbol{s}}_{t|t-1})$$

where $\boldsymbol{M}$ and $\bar{\boldsymbol{M}}$ are the slope and constant matrix of the measurement equation in (4.13) and (4.14). The innovation covariance is

$$\boldsymbol{P}^{yy}_{t|t-1} = \boldsymbol{M}\boldsymbol{P}^{ss}_{t|t-1}\boldsymbol{M}' + \boldsymbol{R}$$

and the log likelihood contribution to the sample log likelihood is

$$l_t = -0.5(d_y \log(2\pi) + \log|\boldsymbol{P}^{yy}_{t|t-1}|) + \boldsymbol{y}^i_t(\boldsymbol{P}^{yy}_{t|t-1})^{-1}(\boldsymbol{y}^i_t)'.$$

The posterior can be calculated by

$$\begin{aligned}p(\boldsymbol{s}_{t|t}) &= \mathcal{N}(\boldsymbol{s}_{t|t}; \hat{\boldsymbol{s}}_{t|t}, \boldsymbol{P}^{ss}_{t|t}) \\ \hat{\boldsymbol{s}}_{t|t} &= \hat{\boldsymbol{s}}_{t|t-1} + \boldsymbol{P}^{ss}_{t|t-1}\boldsymbol{M}'(\boldsymbol{P}^{yy}_{t|t-1})^{-1} \\ \boldsymbol{P}^{ss}_{t|t} &= \boldsymbol{P}^{ss}_{t|t-1} - \boldsymbol{P}^{ss}_{t|t-1}\boldsymbol{M}'(\boldsymbol{P}^{yy}_{t|t-1})^{-1}\boldsymbol{M}\boldsymbol{P}^{ss}_{t|t-1}.\end{aligned}$$

4.4.3.3 Gaussian

This section derives the posterior density for the special case of additive and Gaussian shocks. It serves the purpose to work out where exactly integration is needed and the Smolyak operator can be of use. Since the prior and posterior are conditionally normal only the mean and covariance have to be updated. The state and measurement equations are given by

$$\begin{aligned} \boldsymbol{s}_t &= \boldsymbol{g}(\boldsymbol{s}_{t-1}) + \boldsymbol{e}_t \\ \boldsymbol{y}_t &= \boldsymbol{m}(\boldsymbol{s}_t) + \boldsymbol{\epsilon}_t \end{aligned} \tag{4.35}$$

with normal shock distributions $\boldsymbol{e}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_t)$ and $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_t)$. The prior of the states is given by $\mathcal{N}(\boldsymbol{s}_0; \hat{\boldsymbol{s}}_0, \boldsymbol{P}_0^{\boldsymbol{ss}})$ and the predictive density of equation (4.35) is

$$p\left(\boldsymbol{s}_t \,|\, \boldsymbol{s}_{t-1}\right) = \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{g}(\boldsymbol{s}_{t-1}), \boldsymbol{Q}_t).$$

The prediction step (4.28) of the general filter recursion can be written as

$$p\left(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t-1}\right) = \int \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{g}(\boldsymbol{s}_{t-1}), \boldsymbol{Q}_t)\, p\left(\boldsymbol{s}_{t-1} \,|\, \boldsymbol{y}_{1:t-1}\right) d\boldsymbol{s}_{t-1}. \tag{4.36}$$

Since the expected value of a normally distributed variable is

$$E(\boldsymbol{t}) \equiv \hat{\boldsymbol{t}} = \int \boldsymbol{t}\mathcal{N}(\boldsymbol{t}; \boldsymbol{f}(\boldsymbol{s}), \boldsymbol{\Sigma})\, d\boldsymbol{t} = \boldsymbol{f}(\boldsymbol{s})$$

equation (4.36) can be used to derive the prior state mean

$$\begin{aligned} E(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t-1}) \equiv \hat{\boldsymbol{s}}_{t|t-1} &= \int \boldsymbol{s}_t p\left(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t-1}\right) d\boldsymbol{s}_t \\ &= \int \boldsymbol{s}_t \left(\int \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{g}(\boldsymbol{s}_{t-1}), \boldsymbol{Q}_t)\, p\left(\boldsymbol{s}_{t-1} \,|\, \boldsymbol{y}_{1:t-1}\right) d\boldsymbol{s}_{t-1} \right) d\boldsymbol{s}_t \\ &= \int \left(\int \boldsymbol{s}_t \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{g}(\boldsymbol{s}_{t-1}), \boldsymbol{Q}_t)\, d\boldsymbol{s}_t \right) p\left(\boldsymbol{s}_{t-1} \,|\, \boldsymbol{y}_{1:t-1}\right) d\boldsymbol{s}_{t-1} \\ &= \int \boldsymbol{g}(\boldsymbol{s}_{t-1})\, p\left(\boldsymbol{s}_{t-1} \,|\, \boldsymbol{y}_{1:t-1}\right) d\boldsymbol{s}_{t-1}. \end{aligned}$$

If the previous posterior density is

$$p\left(\boldsymbol{s}_{t-1} \,|\, \boldsymbol{y}_{1:t-1}\right) = \mathcal{N}(\boldsymbol{s}_{t-1}; \hat{\boldsymbol{s}}_{t-1|t-1}, \boldsymbol{P}_{t-1|t-1}^{\boldsymbol{ss}}) \tag{4.37}$$

the prediction density is

$$p\left(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{t-1}\right) = \mathcal{N}(\boldsymbol{s}_t; \hat{\boldsymbol{s}}_{t|t-1}, \boldsymbol{P}_{t|t-1}^{\boldsymbol{ss}})$$

$$\hat{\boldsymbol{s}}_{t|t-1} = \int \boldsymbol{g}(\boldsymbol{s}_{t-1}) \qquad \mathcal{N}(\boldsymbol{s}_{t-1}; \hat{\boldsymbol{s}}_{t-1|t-1}, \boldsymbol{P}_{t-1|t-1}^{\boldsymbol{ss}})\, d\boldsymbol{s}_{t-1} \tag{4.38}$$

$$\begin{aligned} \boldsymbol{P}_{t|t-1}^{\boldsymbol{ss}} &= \int \boldsymbol{g}(\boldsymbol{s}_{t-1})\boldsymbol{g}^T(\boldsymbol{s}_{t-1}) \mathcal{N}(\boldsymbol{s}_{t-1}; \hat{\boldsymbol{s}}_{t-1|t-1}, \boldsymbol{P}_{t-1|t-1}^{\boldsymbol{ss}})\, d\boldsymbol{s}_{t-1} \\ &\quad + \boldsymbol{Q}_t - \hat{\boldsymbol{s}}_{t|t-1}\hat{\boldsymbol{s}}_{t|t-1}^T. \end{aligned} \tag{4.39}$$

The expected value of the observed variable is then given by

$$\begin{aligned}\hat{\boldsymbol{y}}_{t|t-1} &= \int \boldsymbol{y}_t \qquad \mathcal{N}(\boldsymbol{s}_t; \hat{\boldsymbol{s}}_{t|t-1}, \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1})\, d\boldsymbol{s}_t \\ &= \int \boldsymbol{m}(\boldsymbol{s}_t)\mathcal{N}(\boldsymbol{s}_t; \hat{\boldsymbol{s}}_{t|t-1}, \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1})\, d\boldsymbol{s}_t. \end{aligned} \tag{4.40}$$

With $\boldsymbol{\epsilon}^m_{t|t-1} \equiv \boldsymbol{m}(\boldsymbol{s}_t) - \hat{\boldsymbol{y}}_{t|t-1}$ the covariance of $\boldsymbol{y}_t$ is

$$\begin{aligned}\boldsymbol{P}^{\boldsymbol{yy}}_{t|t-1} &= E\left(\left[\boldsymbol{\epsilon}^m_{t|t-1}\right]\left[\boldsymbol{\epsilon}^m_{t|t-1}\right]^T\right) \\ &= \int \boldsymbol{m}(\boldsymbol{s}_t)\,\boldsymbol{m}^T(\boldsymbol{s}_t)\,\mathcal{N}(\boldsymbol{s}_t; \hat{\boldsymbol{s}}_{t|t-1}, \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1})\, d\boldsymbol{s}_t + \boldsymbol{H}_t - \hat{\boldsymbol{y}}_{t|t-1}\hat{\boldsymbol{y}}^T_{t|t-1}. \end{aligned} \tag{4.41}$$

The last section shows that the Kalman filter step is also useful for the nonlinear Gaussian model and we can calculate the covariance

$$\begin{aligned}\boldsymbol{P}^{\boldsymbol{sy}}_{t|t-1} &= E\left(\left[\boldsymbol{s}_t - \hat{\boldsymbol{s}}_{t|t-1}\right]\left[\boldsymbol{\epsilon}^m_{t|t-1}\right]^T\right) \\ &= \int \boldsymbol{s}_t\boldsymbol{m}^T(\boldsymbol{s}_t)\mathcal{N}(\boldsymbol{s}_t; \hat{\boldsymbol{s}}_{t|t-1}, \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1})\, d\boldsymbol{s}_t - \hat{\boldsymbol{s}}_{t|t-1}\hat{\boldsymbol{y}}^T_{t|t-1} \end{aligned} \tag{4.42}$$

in order to obtain the Kalman gain

$$\boldsymbol{K}_t = \boldsymbol{P}^{\boldsymbol{sy}}_{t|t-1}(\boldsymbol{P}^{\boldsymbol{yy}}_{t|t-1})^{-1}.$$

The recursion is closed by the filter step (4.29) and we obtain the next posterior density

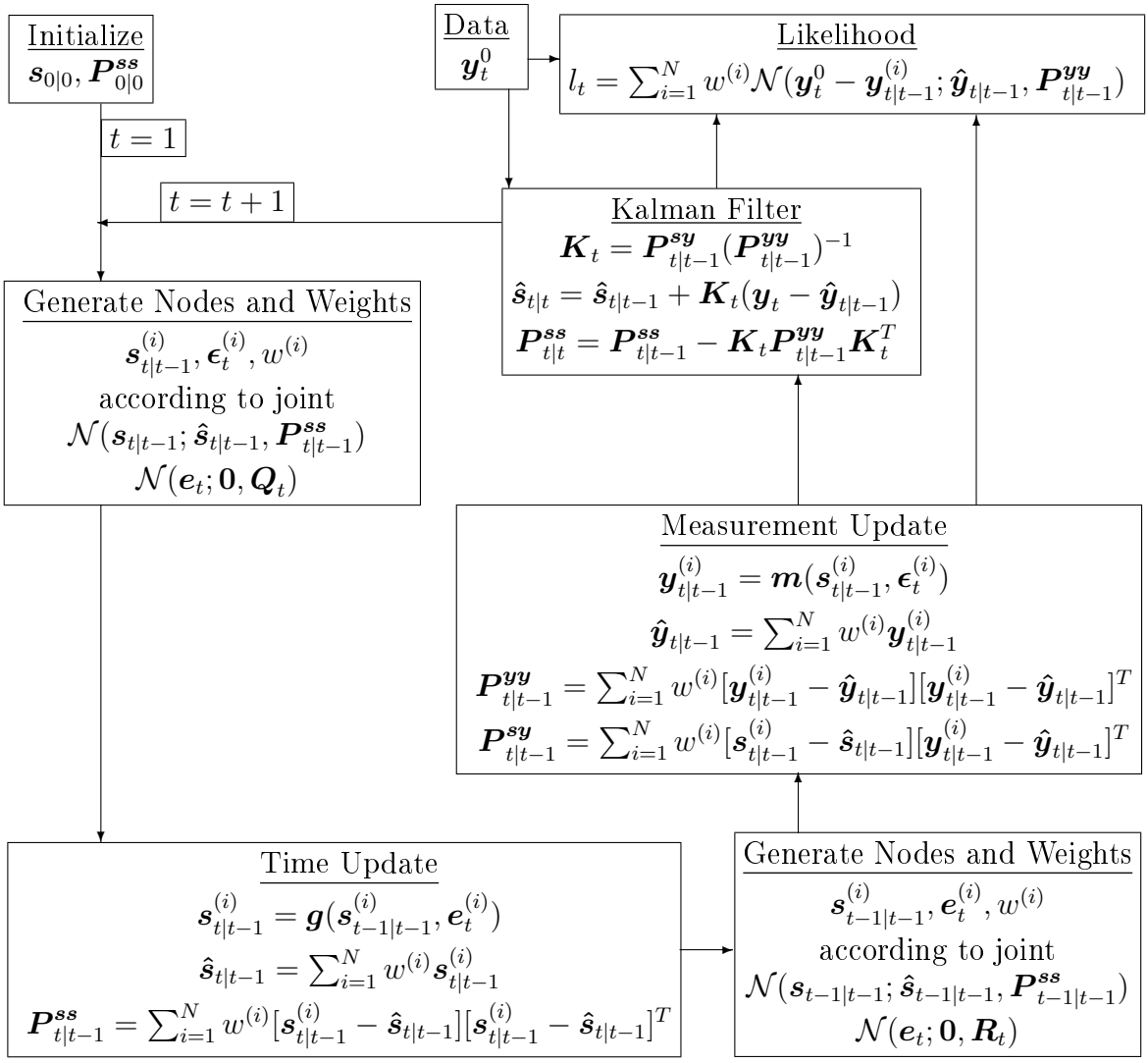
$$\begin{aligned} p\left(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t}\right) &= \mathcal{N}(\boldsymbol{s}_t; \hat{\boldsymbol{s}}_{t|t}, \boldsymbol{P}^{\boldsymbol{ss}}_{t|t}) \\ \hat{\boldsymbol{s}}_{t|t} &= \hat{\boldsymbol{s}}_{t|t-1} + \boldsymbol{K}_t[\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1}] \\ \boldsymbol{P}^{\boldsymbol{ss}}_{t|t} &= \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1} - \boldsymbol{K}_t\boldsymbol{P}^{\boldsymbol{yy}}_{t|t-1}\boldsymbol{K}^T_t. \end{aligned}$$

The mathematical problem and the only approximation needed for this filter is the evaluation of the integrals for the mean and covariance of the state prediction in equations (4.38) and (4.39), the expected value of the observables in equation (4.40), the innovation covariance in equation (4.41) and the covariance between the states and observations in equation (4.42). The integrals involved have the form

$$\begin{aligned}\mathcal{I}(\boldsymbol{f})_{\mathcal{N}} &= \int \boldsymbol{f}(\boldsymbol{s})\,\mathcal{N}(\boldsymbol{s}, \hat{\boldsymbol{s}}, \boldsymbol{P}^{\boldsymbol{ss}})\, d\boldsymbol{s} \\ &\approx \int \boldsymbol{f}(\boldsymbol{s}) \sum_{i=1}^{N} w^{(i)}\delta(\boldsymbol{s} - \boldsymbol{s}^{(i)})\, d\boldsymbol{s} = \sum_{i=1}^{N} w^{(i)}\, \boldsymbol{f}(\boldsymbol{s}^{(i)}) \end{aligned}$$

and can therefore be approximated by Gaussian quadrature by means of nodes $\boldsymbol{s}^{(j)}$ and weights $w^{(j)}$. Figure 4.8 shows the recursive steps for a general non-additive noise model. With additive Gaussian noise only the integrals for two

Figure 4.8: Gaussian Filter



moments have to be approximated. For the general nonadditive model the approximation encompasses also the assumption of normally distributed prior and posterior densities. The integrands in (4.38) to (4.42) have then to be evaluated at the joint nodes and weights for states and shocks.

The approximation of the prediction mean and covariance in equations (4.38), (4.39) for the additive shock model is given by

$$\begin{aligned}\hat{\boldsymbol{s}}_{t|t-1} &= \sum_{i=1}^{N} w^{(i)} \boldsymbol{g}(\boldsymbol{s}_{t-1|t-1}^{(i)}) \\ \boldsymbol{P}_{t|t-1}^{\boldsymbol{ss}} &= \sum_{i=1}^{N} w^{(i)} \boldsymbol{g}(\boldsymbol{s}_{t-1|t-1}^{(i)}) \boldsymbol{g}^T(\boldsymbol{s}_{t-1|t-1}^{(i)}) + \boldsymbol{Q}_t - \hat{\boldsymbol{s}}_{t|t-1} \hat{\boldsymbol{s}}_{t|t-1}^T.\end{aligned}$$

From the density $\mathcal{N}(\boldsymbol{s}_{t|t-1}; \hat{\boldsymbol{s}}_{t|t-1}, \boldsymbol{P}_{t|t-1}^{\boldsymbol{ss}})$ new nodes and weights are generated and the measurement update moments in equations (4.40), (4.41), (4.42) are calculated by

$$\begin{aligned}\hat{\boldsymbol{y}}_{t|t-1} &= \sum_{i=1}^{N} w^{(i)} \boldsymbol{m}(\boldsymbol{s}_{t|t-1}^{(i)}) \\ \boldsymbol{P}_{t|t-1}^{\boldsymbol{yy}} &= \sum_{i=1}^{N} w^{(i)} \boldsymbol{m}(\boldsymbol{s}_{t|t-1}^{(i)}) \boldsymbol{m}^T(\boldsymbol{s}_{t|t-1}^{(i)}) + \boldsymbol{R} - \hat{\boldsymbol{y}}_{t|t-1} \hat{\boldsymbol{y}}_{t|t-1}^T \\ \boldsymbol{P}_{t|t-1}^{\boldsymbol{sy}} &= \sum_{i=1}^{N} w^{(i)} [\boldsymbol{s}_{t|t-1}^{(i)} - \hat{\boldsymbol{s}}_{t|t-1}][\boldsymbol{m}(\boldsymbol{s}_{t|t-1}^{(i)}) - \hat{\boldsymbol{y}}_{t|t-1}]^T.\end{aligned}$$

4.4.4 Particle Filter

The last section assumed that the posterior density $p(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t})$ is Gaussian. It allows to generate nodes and weights and to update this discrete density approximation. For general nonlinear state space models and possibly non-Gaussian shocks the posterior is nonstandard or even multimodal. Without an analytical expression for these densities it is impossible to directly generate nodes and weights or random draws. Nevertheless there are indirect ways to generate random draws and the one usually used is importance sampling. Another random number generator class are Markov Chain Monte Carlo methods. They do not rely like importance sampling on a proposal density which should be similar to the target density. The price for this generality is a slow convergence towards a representative sample which makes them impractical for recursive filtering. Fortunately good proposal densities for importance sampling filtering can be generated.

The particle filter owes its name to the random draws generated from the posterior density we are interested in. Since we cannot draw particles from the posterior density $p(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t})$ directly we divide it by the proposal density $q(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t})$ and obtain the importance weights

$$w(\boldsymbol{s}_t) = \frac{p(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t})}{q(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t})}.$$

This allows to rewrite the general integral of equation (4.31) as

$$E(\boldsymbol{g}(\boldsymbol{s}_t)) \equiv \hat{\boldsymbol{g}}(\boldsymbol{s}_t) = \frac{\int \boldsymbol{g}(\boldsymbol{s}_t) w(\boldsymbol{s}_t) q(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t})\, d\boldsymbol{s}_t}{\int w(\boldsymbol{s}_t) q(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t})\, d\boldsymbol{s}_t}.$$

N particles drawn from the proposal density $\boldsymbol{s}_t^{(i)} \sim q(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t})$ and the associated weights represent an approximation to the posterior. The integral of interest can be calculated as

$$\hat{\boldsymbol{g}}(\boldsymbol{s}_t) = \frac{\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g}(\boldsymbol{s}_t^{(i)}) w(\boldsymbol{s}_t^{(i)})}{\frac{1}{N}\sum_{i=1}^{N} w(\boldsymbol{s}_t^{(i)})} = \sum_{i=1}^{N} \boldsymbol{g}(\boldsymbol{s}_t^{(i)}) \tilde{w}(\boldsymbol{s}_t^{(i)})$$

with normalized weights

$$\tilde{w}(\boldsymbol{s}_t^{(i)}) = \frac{w(\boldsymbol{s}_t^{(i)})}{\sum_{i=1}^{N} w(\boldsymbol{s}_t^{(i)})}.$$

According to the factorization in equations (4.29) and (4.28) for p and analogous for q the weight formula can be rewritten such that the weights are updated recursively

$$w(\boldsymbol{s}_t) = \frac{p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t})}{q(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t})} = \frac{l_t^{-1} p(\boldsymbol{y}_t \mid \boldsymbol{s}_t)\, p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1})}{q(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t})} \tag{4.43}$$

$$= \frac{l_t^{-1} p(\boldsymbol{y}_t \mid \boldsymbol{s}_t) \int p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1})\, p(\boldsymbol{s}_{t-1} \mid \boldsymbol{y}_{1:t-1})\, d\boldsymbol{s}_{t-1}}{\int q(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1}) q(\boldsymbol{s}_{t-1} \mid \boldsymbol{y}_{1:t-1})\, d\boldsymbol{s}_{t-1}}. \tag{4.44}$$

4.4.4.1 Bootstrap

The discrete approximation of $p(\boldsymbol{s}_{t-1} \mid \boldsymbol{y}_{1:t-1})$ by a set of particles $\boldsymbol{s}_{t-1|t-1}^{(i)}$ and weights $w_{t-1}^{(i)}$ is given by

$$p(\boldsymbol{s}_{t-1} \mid \boldsymbol{y}_{1:t-1}) \approx \sum_{i=1}^{N} w_{t-1}^{(i)} \delta(\boldsymbol{s}_{t-1} - \boldsymbol{s}_{t-1|t-1}^{(i)}) \quad \text{with } w_{t-1}^{(i)} = \frac{p(\boldsymbol{s}_{t-1|t-1}^{(i)})}{q(\boldsymbol{s}_{t-1|t-1}^{(i)})}. \tag{4.45}$$

The next period particles $\boldsymbol{s}_{t|t-1}^{(i)}$ are obtained from equation (4.25)

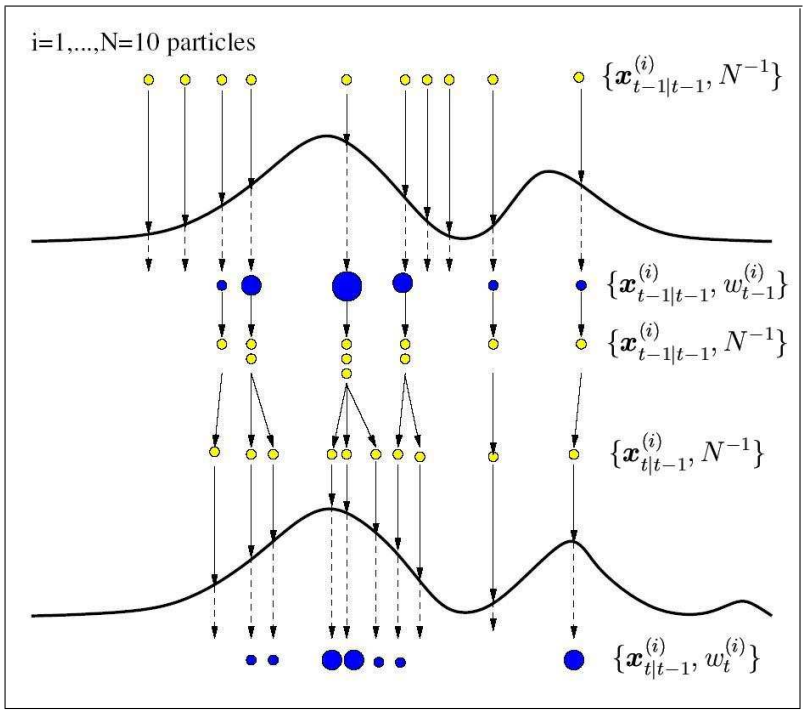
$$\boldsymbol{s}_{t|t-1}^{(i)} = \boldsymbol{g}(\boldsymbol{s}_{t-1|t-1}^{(i)}, \boldsymbol{e}_t^{(i)}) \tag{4.46}$$

where in addition draws from the state shock distribution $\boldsymbol{e}_t^{(i)} \sim q^e(\boldsymbol{e})$ have to be generated. Equation (4.45) allows to write the weight update equation (4.44) as

$$w_t^{(i)} \propto \frac{p(\boldsymbol{y}_t \mid \boldsymbol{s}_{t|t-1}^{(i)})\, p(\boldsymbol{s}_{t|t-1}^{(i)} \mid \boldsymbol{s}_{t-1|t-1}^{(i)})\, p(\boldsymbol{s}_{t-1|t-1}^{(i)})}{q(\boldsymbol{s}_{t|t-1}^{(i)} \mid \boldsymbol{s}_{t-1|t-1}^{(i)}) q(\boldsymbol{s}_{t-1|t-1}^{(i)})}$$

$$= w_{t-1}^{(i)} \frac{p(\boldsymbol{y}_t \mid \boldsymbol{s}_{t|t-1}^{(i)})\, p(\boldsymbol{s}_{t|t-1}^{(i)} \mid \boldsymbol{s}_{t-1|t-1}^{(i)})}{q(\boldsymbol{s}_{t|t-1}^{(i)} \mid \boldsymbol{s}_{t-1|t-1}^{(i)})}.$$

Figure 4.9: Resampling Scheme



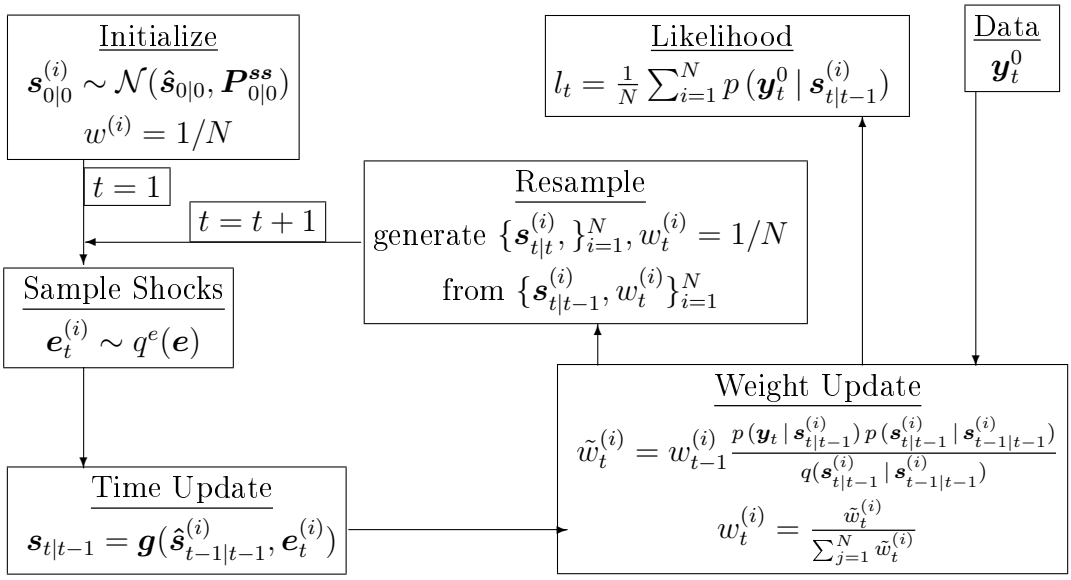
Source: van der Merwe, de Freitas, Doucet, and Wan (2001)

The posterior is finally approximated by the discrete density

$$p\left(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t}\right) \approx \sum_{i=1}^{N} w_t^{(i)} \delta(\boldsymbol{s}_t - \boldsymbol{s}_{t|t}^{(i)}).$$

Where do we get the posterior particles $\boldsymbol{s}_{t|t}^{(i)}$ from? A serious problem with this filter is that a recursive update of the state particles in equation (4.46) increases their variance without bound as $t \to \infty$. The particles move away from the expected value and their weight degenerates to zero. A brute force solution is to use more particles. A more elegant solution was proposed by Gordon, Salmond, and Smith (1993) and resamples the particles at each step according to their weight. The particles with low weights are dropped and the ones with high weights are duplicated. These resampled particles represent an equally weighted posterior sample $\boldsymbol{s}_{t|t}^{(i)}$. This procedure is shown in figure 4.9 and the general particle filter is summarized in figure 4.10.
The implementation of an importance sampling particle filter needs the specification of a proposal density $q(\boldsymbol{s}_{t|t-1} \mid \boldsymbol{s}_{t-1|t-1})$ for the evaluation of probabilities of the importance weights in equation (4.47). Moreover the probabilities $p\left(\boldsymbol{y}_t \mid \boldsymbol{s}_{t|t-1}\right)$ and $p\left(\boldsymbol{s}_{t|t-1} \mid \boldsymbol{s}_{t-1|t-1}\right)$ are needed. They can be difficult to com-

Figure 4.10: Bootstrap Filter



pute if shocks are nonadditive and non-Gaussian and inverse densities have to be evaluated to obtain a likelihood value.
The simplest Monte Carlo variant of a particle filter, called bootstrap or sequential importance resampling particle filter, takes the state transition as the proposal density

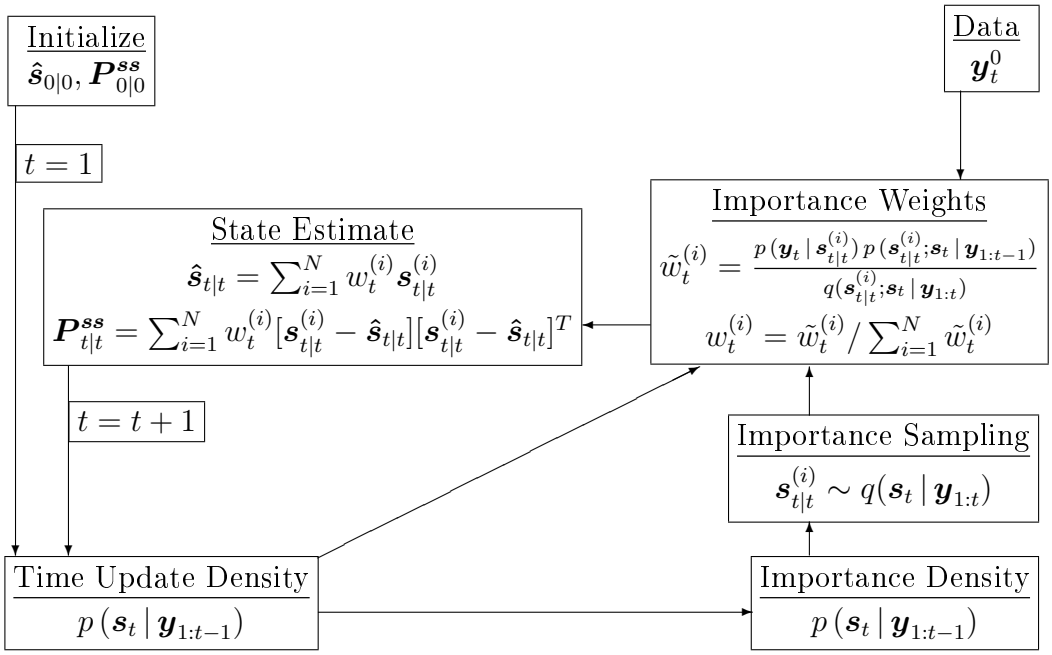
$$q(s_{t|t-1}^{(i)} \mid s_{t-1|t-1}^{(i)}) = p(s_{t|t-1}^{(i)} \mid s_{t-1|t-1}^{(i)}).$$

This simplifies the weight update equation (4.44) and we obtain

$$w_t^{(i)} = w_{t-1}^{(i)} p(y_t \mid s_{t|t-1}^{(i)}).$$

The disadvantage is that the last observation y_t is not taken into account in the importance density to form the posterior. The consequence is that we may sample in low probability regions of the true posterior with low probabilities of the particles. This decreases the effective number of particles and more particles or the resampling step are needed.
The principle of the bootstrap filter is a trial and error approach. Starting from the prior density of the unobservables, one simulates many shock realizations with the given shock variance. This gives the next period state realizations and together with simulated measurement shocks we get many simulated observables. Many of the simulated observables will be very unlikely. These observables and their generating unobservables are dropped for simulations in the following periods. The remaining particles of the unobservables can be taken as an estimate of their density. The bootstrap filter can be interpreted as a genetic algorithm where the resampling step governs the survival of the fittest.

Figure 4.11: General Particle Filter



Fernádez-Villaverde and Rubio-Ramírez (2004b) use this bootstrap filter with 40,000 particles. For each particle and observation the policy for the next period has to be interpolated in the prediction step. This step is a main bottleneck in the whole likelihood approach. In case of a linear finite element approximation this step is a fast table look up to locate the involved subdomain combined with a linear interpolation. In case of spectral approximation the interpolation is more expensive. The price for the reduction of the number of nodes needed for a spectral approximation is therefore an increased effort in the interpolation step within the likelihood evaluation.

4.4.4.2 Gaussian

Particle filters without resampling rely on equation (4.43). For a sample $\boldsymbol{s}^{(i)}_{t|t} \sim q(\boldsymbol{s}_{t|t}, \boldsymbol{y}_{1:t})$ from the importance density the importance weights are

$$w^{(i)}_t \propto \tilde{w}^{(i)}_t = \frac{p(\boldsymbol{y}_t \mid \boldsymbol{s}^{(i)}_{t|t})\, p(\boldsymbol{s}^{(i)}_{t|t}; \boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1})}{q(\boldsymbol{s}^{(i)}_{t|t}; \boldsymbol{s}_t \mid \boldsymbol{y}_{1:t})} \qquad w^{(i)}_t = \frac{\tilde{w}^{(i)}_t}{\sum_{i=1}^N \tilde{w}^{(i)}_t}$$

where $p(x^{(i)}; x \mid y)$ is the density of x evaluated at $x^{(i)}$. This general particle filter is shown in figure 4.11. The Gaussian particle filter by Kotecha and Djurić (2003a)

starts with an approximation to the previous period posterior by a Gaussian density

$$p\left(\boldsymbol{s}_{t-1} \mid \boldsymbol{y}_{1:t-1}\right) \approx \mathcal{N}(\boldsymbol{s}_{t-1}; \hat{\boldsymbol{s}}_{t-1|t-1}, \boldsymbol{P}^{\boldsymbol{ss}}_{t-1|t-1})$$

and generates nodes $\boldsymbol{s}^{(i)}_{t-1|t-1}$ and weights w_i according to this law. The nodes are updated to $\boldsymbol{s}^{(i)}_{t|t-1}$ by the state equation (4.46) and the prior density is then approximated by

$$\begin{aligned} p\left(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1}\right) &\approx \mathcal{N}(\boldsymbol{s}_t; \hat{\boldsymbol{s}}_{t|t-1}, \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1}) \\ \hat{\boldsymbol{s}}_{t|t-1} &= \sum_{i=1}^{N} w_i \boldsymbol{s}^{(i)}_{t|t-1} \\ \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1} &= \sum_{i=1}^{N} w_i [\boldsymbol{s}^{(i)}_{t|t-1} - \hat{\boldsymbol{s}}_{t|t-1}][\boldsymbol{s}^{(i)}_{t|t-1} - \hat{\boldsymbol{s}}_{t|t-1}]^T. \end{aligned}$$

Particles $\boldsymbol{s}^{(i)}_{t|t}$ are drawn from the importance density and by

$$w^{(i)}_t \propto \frac{p\left(\boldsymbol{y}_t \mid \boldsymbol{s}^{(i)}_{t|t}\right)\mathcal{N}(\boldsymbol{s}^{(i)}_{t|t}; \hat{\boldsymbol{s}}_{t|t-1}, \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1})}{q(\boldsymbol{s}^{(i)}_{t|t}; \boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1})}$$

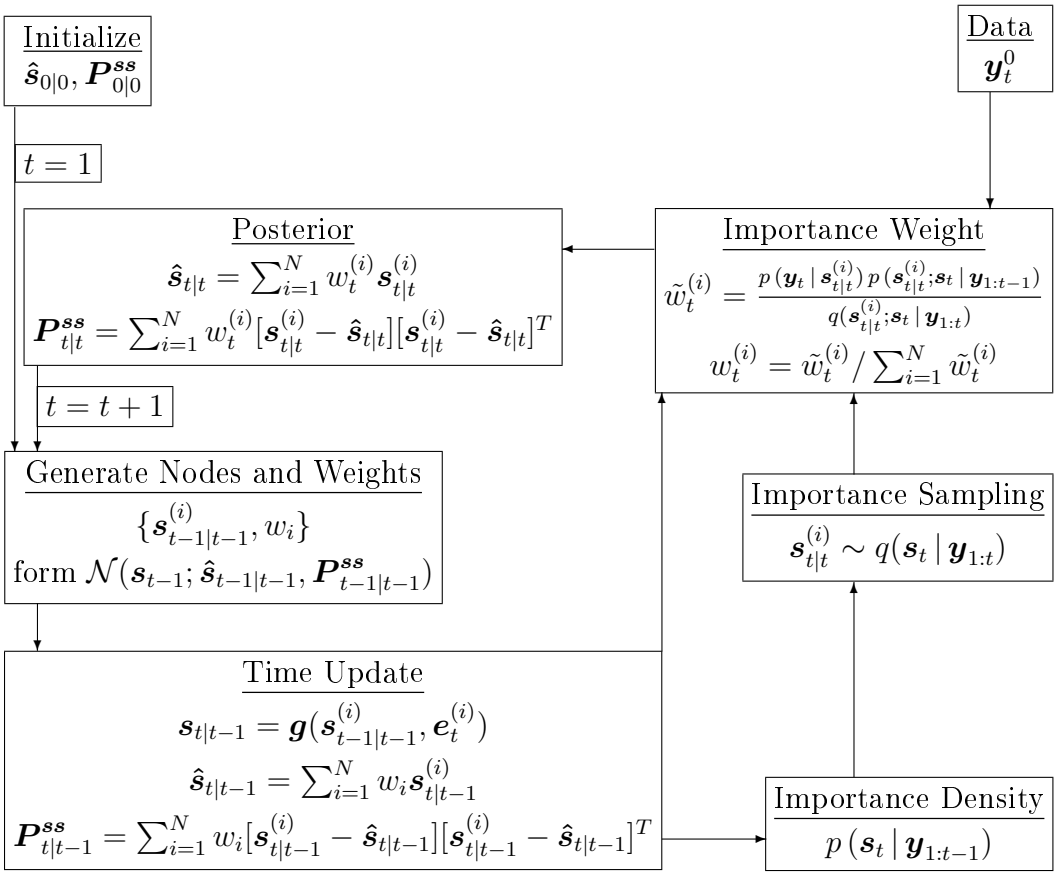
their weights are calculated. Finally the approximation of the posterior is calculated by

$$\begin{aligned} p\left(\boldsymbol{s}_{t|t} \mid \boldsymbol{y}_{1:t}\right) &\approx \mathcal{N}(\boldsymbol{s}_t; \hat{\boldsymbol{s}}_{t|t}, \boldsymbol{P}^{\boldsymbol{ss}}_{t|t}) \\ \hat{\boldsymbol{s}}_{t|t} &= \sum_{i=1}^{N} w^{(i)}_t \boldsymbol{s}^{(i)}_{t|t} \\ \boldsymbol{P}^{\boldsymbol{ss}}_{t|t} &= \sum_{i=1}^{N} w^{(i)}_t [\boldsymbol{s}^{(i)}_{t|t} - \hat{\boldsymbol{s}}_{t|t}][\boldsymbol{s}^{(i)}_{t|t} - \hat{\boldsymbol{s}}_{t|t}]^T. \end{aligned}$$

This filter is the basis of a more general filter provided in Kotecha and Djurić (2003b) as a companion paper. It is a generalization towards a more accurate approximation of the posterior compared to a simple Gaussian density. The approximation is constructed as a sum of Gaussian densities to approximate more than just the first two moments of the posteriors. The moment update in the basic Gaussian particle filter is shown in figure 4.12. The more general filter is yet not implemented but it could be an interesting next step after the Gaussian filters are tested.

The Gaussian particle filter still lacks the specification of an importance density. One approach is to combine it with one of the nonlinear Gaussian filters. van der Merwe, de Freitas, Doucet, and Wan (2001) combined the unscented Kalman

Figure 4.12: Gaussian Particle



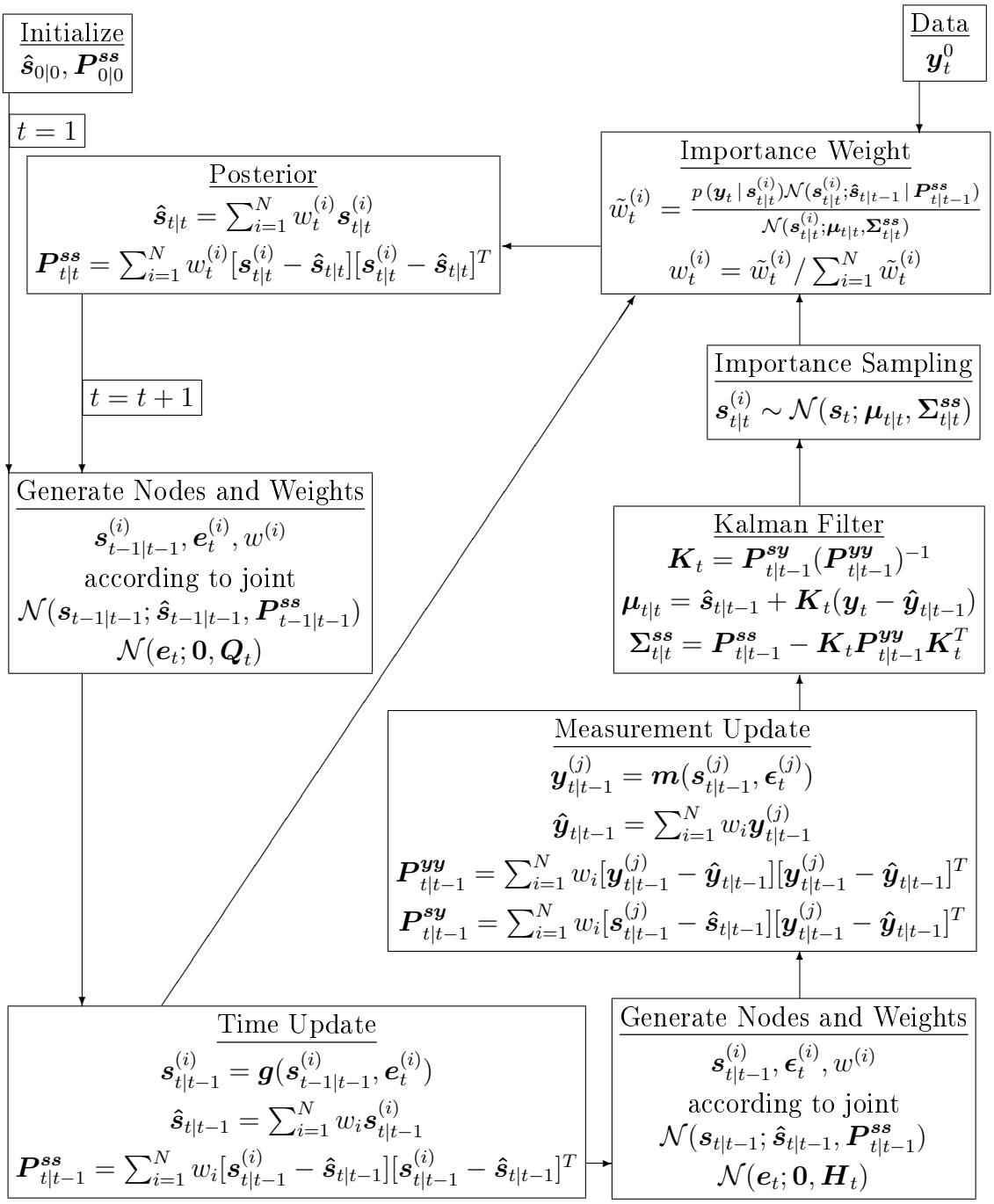
filter with the bootstrap filter and improved the performance substantially. I implemented the Gaussian particle filter where the Gaussian filter provides the importance density.
As in the Gaussian filter the prior is approximated by a normal density

$$p(\boldsymbol{s}_t \mid \boldsymbol{y}_{1:t-1}) \approx \mathcal{N}(\boldsymbol{s}_t; \hat{\boldsymbol{s}}_{t|t-1}, \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1})$$

and the prior particles and their moments $\hat{\boldsymbol{s}}_{t|t-1}, \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1}$ are calculated by the time update

$$\begin{aligned}
\boldsymbol{s}_{t|t-1} &= \boldsymbol{g}(\boldsymbol{s}_{t-1|t-1}^{(i)}, \boldsymbol{e}_t^{(i)}) \\
\hat{\boldsymbol{s}}_{t|t-1} &= \sum_{i=1}^N w_{t-1}^{(i)} \boldsymbol{s}_{t|t-1}^{(i)} \\
\boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1} &= \sum_{i=1}^N w_{t-1}^{(i)} [\boldsymbol{s}_{t|t-1}^{(i)} - \hat{\boldsymbol{s}}_{t|t-1}][\boldsymbol{s}_{t|t-1}^{(i)} - \hat{\boldsymbol{s}}_{t|t-1}]^T.
\end{aligned}$$

Figure 4.13: Gaussian Particle Filter



For this density nodes $\boldsymbol{s}^{(j)}_{t|t-1}$ and weights $w^{(j)}$ are generated

$\boldsymbol{s}^{(j)}_{t|t-1} \sim \mathcal{N}(\boldsymbol{s}_t; \hat{\boldsymbol{s}}_{t|t-1}, \boldsymbol{P}^{\boldsymbol{ss}}_{t|t-1})$

and used for the measurement update

$$\begin{aligned}
\boldsymbol{y}_{t|t-1}^{(j)} &= \boldsymbol{m}(\boldsymbol{s}_{t|t-1}^{(j)}, \boldsymbol{\epsilon}_t^{(j)}) \\
\hat{\boldsymbol{y}}_{t|t-1} &= \sum_{i=1}^{N} w_j \boldsymbol{y}_{t|t-1}^{(j)}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{P}_{t|t-1}^{\boldsymbol{yy}} &= \sum_{i=1}^{N} w_j [\boldsymbol{y}_{t|t-1}^{(j)} - \hat{\boldsymbol{y}}_{t|t-1}][\boldsymbol{y}_{t|t-1}^{(j)} - \hat{\boldsymbol{y}}_{t|t-1}]^T \\
\boldsymbol{P}_{t|t-1}^{\boldsymbol{sy}} &= \sum_{i=1}^{N} w_j [\boldsymbol{s}_{t|t-1}^{(j)} - \hat{\boldsymbol{s}}_{t|t-1}][\boldsymbol{y}_{t|t-1}^{(j)} - \hat{\boldsymbol{y}}_{t|t-1}]^T
\end{aligned}$$

to obtain the importance density via the Kalman gain

$$\begin{aligned}
\boldsymbol{K}_t &= \boldsymbol{P}_{t|t-1}^{\boldsymbol{sy}} (\boldsymbol{P}_{t|t-1}^{\boldsymbol{yy}})^{-1} \\
q(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t}) &= \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}^{\boldsymbol{ss}}) \\
\boldsymbol{\mu}_{t|t} &= \hat{\boldsymbol{s}}_{t|t-1} + \boldsymbol{K}_t (\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1}) \\
\boldsymbol{\Sigma}_{t|t}^{\boldsymbol{ss}} &= \boldsymbol{P}_{t|t-1}^{\boldsymbol{ss}} - \boldsymbol{K}_t \boldsymbol{P}_{t|t-1}^{\boldsymbol{yy}} \boldsymbol{K}_t^T.
\end{aligned}$$

The importance density is used to generate particles

$$\boldsymbol{s}_{t|t}^{(i)} \sim \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}^{\boldsymbol{ss}})$$

and to calculate the associate importance weights

$$\begin{aligned}
\tilde{w}_t^{(i)} &= \frac{p\,(\boldsymbol{y}_t \,|\, \boldsymbol{s}_{t|t}^{(i)}) \mathcal{N}(\boldsymbol{s}_t; \hat{\boldsymbol{s}}_{t|t-1}, \boldsymbol{P}_{t|t-1}^{\boldsymbol{ss}})}{\mathcal{N}(\boldsymbol{s}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}^{\boldsymbol{ss}})} \\
w_t^{(i)} &= \tilde{w}_t^{(i)} \sum_{i=1}^{N} \tilde{w}_t^{(i)}
\end{aligned}$$

which determine the posterior through its moments

$$\begin{aligned}
p\,(\boldsymbol{s}_t \,|\, \boldsymbol{y}_{1:t}) &\approx \mathcal{N}(\boldsymbol{s}_t; \hat{\boldsymbol{s}}_{t|t}, \boldsymbol{P}_{t|t}^{\boldsymbol{ss}}) \\
\hat{\boldsymbol{s}}_{t|t} &= \sum_{i=1}^{N} w_t^{(i)} \boldsymbol{s}_{t|t}^{(i)} \\
\boldsymbol{P}_{t|t}^{\boldsymbol{ss}} &= \sum_{i=1}^{N} w_t^{(i)} [\boldsymbol{s}_{t|t}^{(i)} - \hat{\boldsymbol{s}}_{t|t}][\boldsymbol{s}_{t|t}^{(i)} - \hat{\boldsymbol{s}}_{t|t}]^T.
\end{aligned}$$

This filter is summarized in figure 4.13.

4.5 Posterior Density

In the Bayesian framework information accumulation is described by the Bayes formula. The basis for inference and object of interest is the posterior of the unobservables

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\int p(\boldsymbol{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}} \propto p(\boldsymbol{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta}).$$

It represents information about the unobservables $\boldsymbol{\theta}$ after available data is processed. $p(\boldsymbol{y})$ is the marginal likelihood and at the heart of the Bayesian model selection in section 4.6. The researcher's or the application specific utility function determines a point estimate from the posterior density. This is a feature which is not automatically embedded in the classical approach although the optimal point estimator depends on the utility function and risk aversion in a given application. An analytical expression is neither available for the likelihood nor for the posterior. The object of interest is a density and an approximation can be obtained by a random number generator.

4.5.1 Metropolis-Hastings

The Metropolis-Hastings algorithm is a subspecies of a Markov Chain Monte Carlo algorithm and allows to generate draws from any target density $p(\boldsymbol{x})$. As opposed to importance sampling no proposal density is needed. The only prerequisite is that the target density can be evaluated at any point $\boldsymbol{x}$ of its domain. The term Markov Chain refers to the fact that draws from the target density are Markovian and not independent. The algorithm is constructed as described in Chib and Greenberg (1995) in such a way that the density of the sequence of draws is the density of interest.
This algorithm is used to draw a sequence $\{\hat{\boldsymbol{\theta}}_n\}_{n=1}^{N}$ of structural parameter vectors from its posterior density so that for large N the sequence is distributed according to the posterior. It is then approximated by a histogram. The algorithm is summarized in table 4.3.
We start with a vector of structural parameters $\hat{\boldsymbol{\theta}}_0$ and draw a candidate vector $\hat{\boldsymbol{\theta}}_n^*$. In the basic Metropolis-Hastings algorithm a candidate is generated by a random walk with $\mathcal{N}(\hat{\boldsymbol{\theta}}_n^*; \hat{\boldsymbol{\theta}}_{n-1}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$. Vector $\hat{\boldsymbol{\theta}}_0$ is the first member of the sequence of draws from the posterior. For $\hat{\boldsymbol{\theta}}_0$ and the candidate vectors the posterior kernel is calculated by evaluating the prior and likelihood. In step 2 (b) the ratio of these two posterior values is calculated. The candidate parameter vector is accepted as the second member of the sequence if the ratio of candidate to the last vector posterior is higher than a uniformly distributed random number between zero and one. The chance of the candidate vector to be accepted increases with the posterior ratio. If the candidate posterior value is higher than the last value the parameter vector will be accepted for sure since the uniform random number

Table 4.3: Metropolis-Hastings Algorithm

1. choose $\hat{\boldsymbol{\theta}}_0$, N and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ such that acceptance ratio is $\approx 30\%$
2. repeat the following steps starting with $n = 1$
   (a) generate $\hat{\boldsymbol{\theta}}_n^* = \hat{\boldsymbol{\theta}}_{n-1} + \boldsymbol{\epsilon}$, where $\mathcal{N}(\boldsymbol{\epsilon}; \mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$
   (b) $\hat{\boldsymbol{\theta}}_n = \begin{cases} \hat{\boldsymbol{\theta}}_n^* & \text{if } U(0,1) \leq \frac{p(\boldsymbol{y}_{1:t}^0 \mid \hat{\boldsymbol{\theta}}_n^*)\, p(\hat{\boldsymbol{\theta}}_n^*)}{p(\boldsymbol{y}_{1:t}^0 \mid \hat{\boldsymbol{\theta}}_{n-1})\, p(\hat{\boldsymbol{\theta}}_{n-1})} \\ \hat{\hat{\boldsymbol{\theta}}}_{n-1} & \text{otherwise} \end{cases}$
   (c) calculate diagnostic test, choose J
   (d) if $n < N$, $n = n + 1$, goto 2(a)
3. disregard burn-in draws $\hat{\boldsymbol{\theta}}_1, ..., \hat{\boldsymbol{\theta}}_J$

is one at maximum. But even for lower posterior ratios there is a chance for the candidate to be accepted since the random number can be even lower than the ratio. If the candidate vector is not accepted then the new draw from the posterior density is taken to be the last parameter vector. Again a candidate vector is constructed by a random walk step. This procedure and an acceptance ratio around 0.3 ensure that the sequence of accepted and repeated parameter vectors is distributed according to the true posterior. The acceptance ratio is the ratio of accepted candidates to all generated candidates.
The Metropolis-Hastings algorithm is also a global maximization procedure since it walks through the feasible parameter space and each step is guided by the relative fit in terms of the posterior value. If the prior is flat for all parameters the posterior is proportional to the likelihood and the Metropolis-Hastings algorithm will find the maximum of the likelihood.
The critical choices of the algorithm are the density to generate candidates $\hat{\boldsymbol{\theta}}_n^*$, the starting value $\hat{\boldsymbol{\theta}}_0$ and the number of draws N. The choice of $\hat{\boldsymbol{\theta}}_0$ determines the number of draws needed before convergence of the sequence is detected. The start value might be very far from a representative draw of the target density and many draws are needed to get into a representative region. But when is it representative or equivalently how long should be the burn-in sequence? This will be discussed in the next section about convergence diagnostic tests.
The distributional choice is often a random walk with normal shocks. For a normal target density the optimal choice of the innovation variance is $\boldsymbol{\Sigma}_\epsilon = \text{Cov}(\boldsymbol{x})$. It has to be scaled so that the acceptance ratio is around 0.3. For a normal target density this is achieved by $\gamma^{RW} = 2.38/\sqrt{D}$. Of course the target density and its covariance $\text{Cov}(\boldsymbol{x})$ are not known as it is the object of

interest of the algorithm. In practice only the diagonal of the matrix $\mathbf{\Sigma}_\epsilon$ is tuned. This variance choice influences the region covered by the sequences. Sampling around the mode of the posterior with large variances will generate candidates far from the current value and therefore a low acceptance probability. Smaller variances increases the acceptance ratio with a too small region being covered so that low probability regions are undersampled. The recommended acceptance ratio results from the attempt to balance this trade off. Both a too high and too low variances will end up in high and slowly decreasing autocorrelations of the individual parameter sequences. A convergence test is therefore an important part of the analysis.

4.5.2 Convergence Test

To detect convergence one can either check several parallel sequences or divide one sequence into subsequences. Examining only one subdivided sequence will result in overly optimistic diagnostic tests. Gelman and Rubin (1992) pointed out that lack of convergence, in many problems, can easily be detected from many but not from one sequence.
Either one sequence is divided into two sequences or several sequences are generated from different start values, in both cases the diagnostic test is calculated for a three dimensional tensor $\hat{\boldsymbol{\theta}}$ of size $N \times D \times M$ with elements $\hat{\theta}^d_{n,m}$. D is the number of parameters, N the number of draws and M the number of sequences. $\hat{\boldsymbol{\theta}}_{n,m}$ is a $1 \times D$ vector and represents the n^{th} draw in the m^{th} sequence and $\hat{\boldsymbol{\theta}}_{:,m}$ is a $N \times D$ matrix and represents all draws in sequence m.
Brooks and Gelman (1998) proposed the multivariate potential scale reduction factor R as a diagnostic test. The general idea is to inspect within and between sequence variances and diagnose convergence if they are close to each other. The within sequence variance is the $D \times D$ matrix

$$\boldsymbol{W} = \frac{1}{M(N-1)} \sum_{m=1}^{M} \sum_{n=1}^{N} (\hat{\boldsymbol{\theta}}_{n,m} - \bar{\boldsymbol{\theta}}_m)'(\hat{\boldsymbol{\theta}}_{n,m} - \bar{\boldsymbol{\theta}}_m)$$

where $\bar{\boldsymbol{\theta}}_m = \frac{1}{N} \sum_{n=1}^{N} \hat{\boldsymbol{\theta}}_{n,m}$ is the $1 \times D$ mean vector in sequence m. $\boldsymbol{W}$ is the mean of the variances in each sequence. The between sequence variance $\boldsymbol{B}/N$ is the $D \times D$ matrix

$$\boldsymbol{B}/N = \frac{1}{M-1} \sum_{m=1}^{M} (\bar{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}})'(\bar{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}})$$

where $\bar{\boldsymbol{\theta}} = \frac{1}{M} \sum_{m=1}^{M} \bar{\boldsymbol{\theta}}_m$ is the $1 \times D$ mean of all draws. The combined variance can be estimated as

$$\boldsymbol{V} = \frac{N-1}{N} \boldsymbol{W} + (1 + \frac{1}{M}) \boldsymbol{B}/N.$$

Convergence is detected for similar $\boldsymbol{V}$ and $\boldsymbol{W}$. A distance measure represents the multivariate potential scale reduction factor

$$R = \frac{N-1}{N} + \frac{M+1}{M}\lambda_{max} \qquad \text{where } \lambda_{max} = \max_{\boldsymbol{a}} \frac{\boldsymbol{a}'\boldsymbol{V}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{W}\boldsymbol{a}}.$$

λ_{max} can be calculated by taking the largest absolute eigenvalue of $\boldsymbol{W}^{-1}\boldsymbol{B}/N$. There are three conditions for convergence

- $\boldsymbol{V}$ and $\boldsymbol{W}$ should stabilize as a function of n,
- R should be below 1.1.

I calculate these numbers repeatedly after some draws and once the conditions are met the burn-in sequence length J is found. The draws thereafter are taken to represent draws from the posterior of structural parameters. My experience so far is that the first two criteria are met before the third. In the estimations I therefore generate burn-in draws until R is below 1.1.

4.5.3 Genetic Extension

The variances on the diagonal of matrix $\gamma^{RW}\boldsymbol{\Sigma}_{\epsilon}$ for the random walk innovations have to be chosen so that an acceptance ratio of around 0.3 is obtained. In the model at hand there are 10 parameters to be estimated and therefore 10 diagonal have to be tuned. To find good values simultaneously is quite demanding and many draws in several trial sequences have to be generated. The necessary number of draws for these tuning runs can be easily as high as for the estimation itself. My experience with the random walk algorithm is that it is quite impossible to tune the covariance to make all parameter sequences simultaneously look like they should. How should they look like?
Figure 4.14 shows three sequences and the associated autocorrelations for one parameter. The upper figure is the consequence of a too high variance. There are too few candidate draws being accepted. The lower graph shows a sequence with a too small variance. Both choices result in a slowly decreasing autocorrelation function as can be seen in the right graphs. The middle graph is a sequence with an appropriate variance choice. This eyeball test is the first hint of a wrong choice and after some experience eyeballing comes close to a calculated autocorrelation function.
Fernández-Villaverde and Rubio-Ramírez (2004a) check robustness by running several sequences with different start vectors. If these sequences and the trial runs to detect an appropriate innovation variance are run simultaneously and not sequentially, then one can assure robustness with respect to start values, calculate unbiased convergence diagnostic tests and in the proposed variant automatically arrives at an appropriate choice of the random walk shock variances.

Figure 4.14: Random Walk Covariance Choice

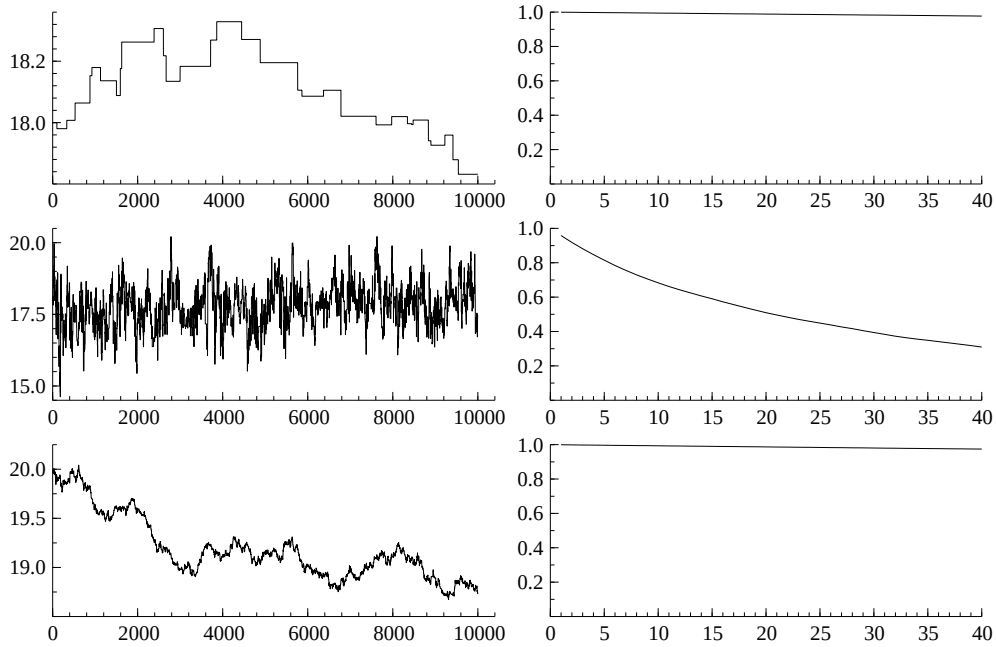


Table 4.4: Genetic Metropolis-Hastings Algorithm

1. choose $\hat{\boldsymbol{\theta}}_m$, $m = 1, ..., M$ and N
2. repeat the following steps starting with $n = 1$
   (a) repeat for $m = 1, ..., M$
      i. draw m_1 and m_2 such that $m_1 \neq m_2 \neq m$
      ii. generate $\hat{\boldsymbol{\theta}}_m^* = \hat{\boldsymbol{\theta}}_m + \gamma^{GE}(\hat{\boldsymbol{\theta}}_{m_1} - \hat{\boldsymbol{\theta}}_{m_2}) + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{U}(-b, b)^D$
      iii. $\hat{\boldsymbol{\theta}}_m = \hat{\boldsymbol{\theta}}_m^*$ if $U(0,1) \leq \frac{p(\boldsymbol{y}_{1:t}^0 \mid \hat{\boldsymbol{\theta}}_i^*)\, p(\hat{\boldsymbol{\theta}}_i^*)}{p(\boldsymbol{y}_{1:t}^0 \mid \hat{\boldsymbol{\theta}}_i)\, p(\hat{\boldsymbol{\theta}}_i)}$
      iv. record $\hat{\boldsymbol{\theta}}_{n,m} = \hat{\boldsymbol{\theta}}_m$
   (b) calculate diagnostic test, set $J = n$ if $R(\hat{\boldsymbol{\theta}}_{1:n,1:M}) < 1.1$
   (c) if $n < N$, $n = n + 1$, goto 2 (a)
3. disregard burn-in draws $\hat{\boldsymbol{\theta}}_{1:J,1:M}$

The draws are collected again in a $N \times D \times M$ tensor $\hat{\boldsymbol{\theta}}$ with M sequences of N draws for D parameters. A candidate draw n^* for parameter vector d in sequence m_i is partly generated as in the random walk variant by a random shock added to the last parameter draw. Moreover, and this is new, the difference between two parameter vectors from two randomly chosen sequences m_1 and m_2 is added.

$$\hat{\boldsymbol{\theta}}_{n^*,m_i} = \hat{\boldsymbol{\theta}}_{n,m_i} + \gamma^{GE}(\hat{\boldsymbol{\theta}}_{n,m_1} - \hat{\boldsymbol{\theta}}_{n,m_2}) + \boldsymbol{\epsilon} \tag{4.47}$$

where $\boldsymbol{\epsilon} \sim U(-b, b)^D$ and $\gamma^{GE} = 2.38/\sqrt{2D}$ for a normal target density. I also tested zero mean normally distributed shocks without changing the results. Parameters γ^{GE} and the shock bounds determine the relative weight of cross and random innovations.

If the variance of the target density is $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{\theta})$ then the variance of the difference of two population parameter vectors from the sequences m_1 and m_2 is $\mathrm{E}[(\boldsymbol{\theta}_{m_1} - \boldsymbol{\theta}_{m_2})(\boldsymbol{\theta}_{m_1} - \boldsymbol{\theta}_{m_2})'] = 2\boldsymbol{\Sigma}$. In case of a converged sequence we get by the law of large numbers $\lim_{N\to\infty} \sum_{n=1}^{N} (\hat{\boldsymbol{\theta}}_{n,m_1} - \hat{\boldsymbol{\theta}}_{n,m_2})(\hat{\boldsymbol{\theta}}_{n,m_1} - \hat{\boldsymbol{\theta}}_{n,m_2})' = 2\boldsymbol{\Sigma}$. Therefore the optimal scale parameter for a normal distribution in this algorithm is $\gamma^{GE} = \gamma^{RW}/\sqrt{2}$.

The intuition behind this procedure is that the variance of the difference between two randomly drawn parameters is the optimal one given that the sequence has converged. The idea originates from the diagnosis test where the within and between sequence variances are examined and convergence is detected when they are the same. I have found the same construction to generate candidate vectors in a working paper by ter Braak (2004) from a biological institute. He argues that this way to generate candidate vectors is also used in the global genetic maximization algorithm called differential evolution by Storn and Price (1995).

The rest of the algorithm corresponds to the random walk variant. The candidate draw and the last draw are used to calculate the posterior ratio which together with a random number determines the acceptance of the candidate. Instead of D variances as in the random walk variant this algorithm has only two free parameters γ^{GE} and b to be tuned in trial runs.

This genetic extension allows a very simple parallelization of the code for an estimation on a cluster of computers. Each computer generates one sequence and the only information to be communicated between the computers is the $\hat{\boldsymbol{\theta}}_{1:M}$ matrix in step 2(a)iv. Its size is only $D \times M$ so that communication costs are mainly determined by the latency of the network. This procedure exhibits a very favorable linear acceleration since doubling the number of computers also doubles the estimation speed.

Section 4.7 presents a Monte Carlo simulation where the one sequential random walk Metropolis-Hastings is compared to the proposed genetic extension.

4.6 Marginal Likelihood

Model selection is a difficult but important matter and depends usually on a variety of more or less formal criteria and the given application. An frequent and rather informal approach is to examine the models ability to replicate some moments of the data. Another important procedure is to select alternative candidates by their out-of-sample record and Friedman (1953) wrote: '*The only relevant test of validity of a hypothesis is comparison of its predictions with experience.*'
The criterion derived in the following is a general likelihood based criterion and according to Berger and Wolpert (1988) the likelihood contains all relevant information needed. One challenge is that models of interest are often nonnested and do not emerge from each other through parameter restrictions. Consequently classical likelihood ratio tests are not of much help. In practice functional forms, the number of estimated and calibrated parameters or the shock distributions may differ across alternative models.
Another problem is that models are inherently wrong since they are not a true representation but approximations of the reality and are designed to explain some features of the real world in a given application. This is a somewhat delicate statement within a classical approach which adheres to the notion of a true parameter or data generating process.
Landon-Lane (1998) discusses the Bayesian model selection within one-dimensional linear processes. The nonnested nature of alternative models as well as the fact that model are never a true representation of the world can be addressed within a Bayesian framework as described by Fernández-Villaverde and Rubio-Ramírez (2004a). Moreover they address the criticism of the Bayesian model selection to depend on the model priors. They show that asymptotically the best model under the Kullback-Leibler measure will have the highest posterior probability.
For some competing models $\{M_1, ..., M_m\}$, parameter priors and observable densities, the unobservables can be integrated out to arrive at the marginal likelihood

$$p(\boldsymbol{y} \mid M_i) = \int_{\Theta_{M_i}} p(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_i}, M_i)\, p(\boldsymbol{\theta}_{M_i} \mid M_i)\, d\boldsymbol{\theta}_{M_i}. \tag{4.48}$$

The parameter posterior is used for the inference conditional on the adequacy of the model whereas the marginal likelihood is used for a criticism of the entertained model in the light of data.
For any two models (M_i, M_j) and their respective priors $p(M_i)$ and $p(M_j)$ the Bayes formula gives

$$\frac{p(M_i \mid \boldsymbol{y})}{p(M_j \mid \boldsymbol{y})} = \frac{p(M_i)}{p(M_j)} \frac{p(\boldsymbol{y} \mid M_i)}{p(\boldsymbol{y} \mid M_j)}.$$

The expression on the left hand side is the posterior odds ratio. On the right hand side the prior odds ratio is transformed by the Bayes factor. Again the (marginal) likelihood or evidence transforms a prior density into a posterior. A posterior odds ratio greater than 1 favors model M_i and M_j otherwise. There is no confidence interval or significance level for this number.
The most difficult part is to calculate the marginal likelihoods which constitute the Bayes factor. Most of the work to calculate the marginal likelihood is already done once the Metropolis-Hastings algorithm has converged and generated parameter draws from the posterior density and the associated posterior values. Gelfand and Dey (1994) show that with any density $h(\boldsymbol{\theta}_{M_i} \mid M_i)$ we can write

$$\begin{aligned}
& \mathrm{E}_{p(\boldsymbol{\theta}_{M_i} \mid \boldsymbol{y}, M_i)} \left(\frac{h(\boldsymbol{\theta}_{M_i} \mid M_i)}{p(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_i}, M_i)\, p(\boldsymbol{\theta}_{M_i} \mid M_i)} \right) \\
= & \int_{\Theta_{M_i}} \frac{h(\boldsymbol{\theta}_{M_i} \mid M_i)}{p(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_i}, M_i)\, p(\boldsymbol{\theta}_{M_i} \mid M_i)} p(\boldsymbol{\theta}_{M_i} \mid \boldsymbol{y}, M_i)\, d\boldsymbol{\theta}_{M_i} \\
= & \int_{\Theta_{M_i}} \frac{h(\boldsymbol{\theta}_{M_i} \mid M_i)}{p(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_i}, M_i)\, p(\boldsymbol{\theta}_{M_i} \mid M_i)} \frac{p(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_i}, M_i)\, p(\boldsymbol{\theta}_{M_i} \mid M_i)}{\int_{\Theta_{M_i}} p(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_i}, M_i)\, p(\boldsymbol{\theta}_{M_i} \mid M_i)\, d\boldsymbol{\theta}_{M_i}} d\boldsymbol{\theta}_{M_i} \\
= & \frac{\int_{\Theta_{M_i}} h(\boldsymbol{\theta}_{M_i} \mid M_i)\, d\boldsymbol{\theta}_{M_i}}{\int_{\Theta_{M_i}} p(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_i}, M_i)\, p(\boldsymbol{\theta}_{M_i} \mid M_i)\, d\boldsymbol{\theta}_{M_i}} = p(\boldsymbol{y} \mid M_i)^{-1}.
\end{aligned}$$

According to the last equation all we have to do is to calculate a weighted mean of the Metropolis-Hastings sequence. Geweke (1999) proposes the following procedure. Calculate the mean and covariance of the parameter draws for each model M_i

$$\bar{\boldsymbol{\theta}}_{M_i} = \frac{1}{N} \sum_{n=1}^{N} \hat{\boldsymbol{\theta}}_{n,M_i} \qquad \hat{\boldsymbol{\Sigma}}_{M_i} = \frac{1}{N} \sum_{n=1}^{N} (\hat{\boldsymbol{\theta}}_{n,M_i} - \bar{\boldsymbol{\theta}}_{M_i})(\hat{\boldsymbol{\theta}}_{n,M_i} - \bar{\boldsymbol{\theta}}_{M_i})'.$$

If k_{M_i} denotes the number of estimated parameters of a model, define a χ^2 critical value for quantile p

$$\boldsymbol{\Theta}_{M_i} = \left\{ \boldsymbol{\theta} : (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}_{M_i})' \hat{\boldsymbol{\Sigma}}_{M_i}^{-1} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}_{M_i}) \leq \chi^2_{1-p}(k_{M_i}) \right\}.$$

To assure robustness we should examine the results for different quantiles. Geweke (1999) proposes to use quantiles $p = 0.1, ..., 0.9$. Density $h(.)$ is specified to be

$$h(\boldsymbol{\theta}) = p^{-1} (2\pi)^{-\frac{k_{M_i}}{2}} \left| \hat{\boldsymbol{\Sigma}}_{M_i} \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}_{M_i})' \hat{\boldsymbol{\Sigma}}_{M_i}^{-1} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}_{M_i}) \right) I_{\boldsymbol{\Theta}_{M_i}}(\boldsymbol{\theta})$$

with an indicator function $I_S(s) = 1$ if $s \in S$ and 0 otherwise. The estimator of the marginal likelihood is finally given by

$$\hat{p}(\boldsymbol{y} \mid M_i) = \left(\frac{1}{N} \sum_{n=1}^{N} \frac{h(\hat{\boldsymbol{\theta}}_{n,M_i})}{p(\boldsymbol{y}^0 \mid \hat{\boldsymbol{\theta}}_{n,M_i}, M_i)\, p(\hat{\boldsymbol{\theta}}_{n,M_i} \mid M_i)} \right)^{-1}.$$

4.7 Results

In this section the performance of the algorithms is presented. It is divided into two subsections - one for the solution and one for the estimation part. The calibrated parameters for the data simulations are the same as in Fernádez-Villaverde and Rubio-Ramírez (2004b) to allow a direct comparison of the results. Two different parameter calibrations are used. The first scenario is the benchmark case with an almost linear policy function. The second parameter vector represents the nonlinear scenario and implies a curved policy function.
The first topic in the solution part is the approximation accuracy of the linear and nonlinear solutions according to the Euler equation error. The second question is how much the Smolyak operator reduces the required computational effort to achieve a level of accuracy comparable to the approximation with the Kronecker operator.
The second part discusses the likelihood evaluation, the genetic Metropolis-Hastings performance, the parameter estimates and the marginal likelihoods. The likelihood evaluations of the Kalman, Gaussian, Gaussian particle and bootstrap filter are compared. The random walk and the proposed genetic Metropolis-Hastings algorithms are compared in a Monte Carlo simulation study. The parameter estimates are presented for both scenarios and the marginal likelihoods finally select between the nonlinear and the linearized models.

4.7.1 Policy

Table 4.5 shows the bounds of a flat prior and both scenarios represented by different parameter sets. The parametrization differ in risk aversion parameter

Table 4.5: Calibrated parameters

parameters	θ	ρ	τ	α	β	δ	σ_e	σ_{ϵ_y}	σ_{ϵ_l}	σ_{ϵ_i}
benchmark	.357	.95	2	.4	.99	.02	.007	.000158	.0011	.000866
risky	.357	.95	50	.4	.99	.02	.035	.000158	.0011	.000866
upper prior	0	0	0	0	.75	0	0	0	0	0
lower prior	1	1	100	1	1	.05	.1	.01	.01	.01

τ and standard deviation of productivity shock σ_e. The risk aversion combined with large shocks implies a curved policy function incorporating a risk premium. The policy does not include the deterministic steady state and together with the curvature it induces relatively large errors in the solution derived from linearization. The main effect in the risky scenario is that all variables exhibit a higher variance since the driving productivity variance rises from 0.022 to 0.11.
Figure 4.15 shows the policies in the benchmark scenario derived from a linearization and the nonlinear solution. The consumption policies do not differ substantially whereas the labor policy is slightly curved even for the benchmark

Figure 4.15: Policy Functions in the Benchmark Scenario

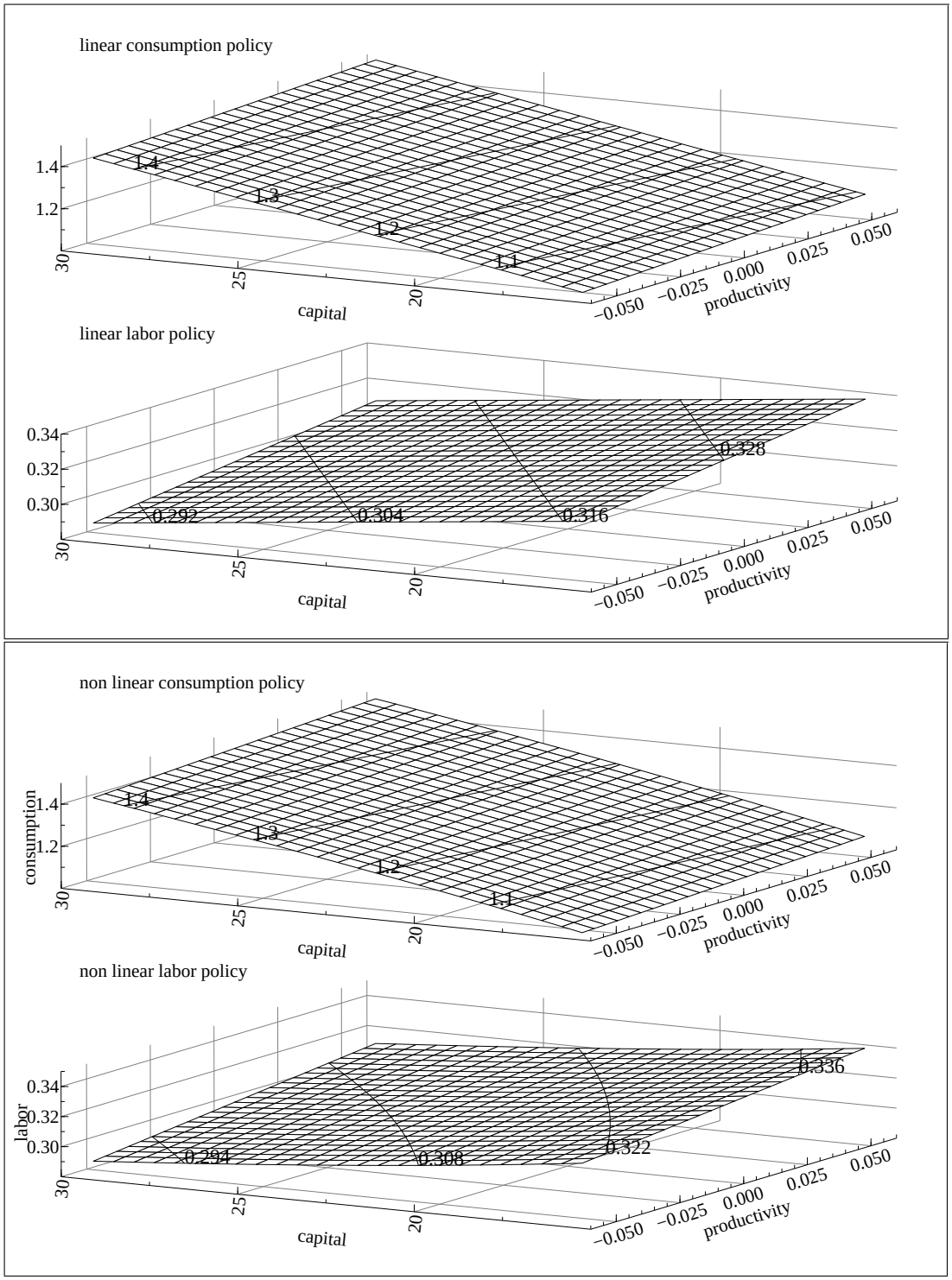


Figure 4.16: Euler Error in the Benchmark Scenario

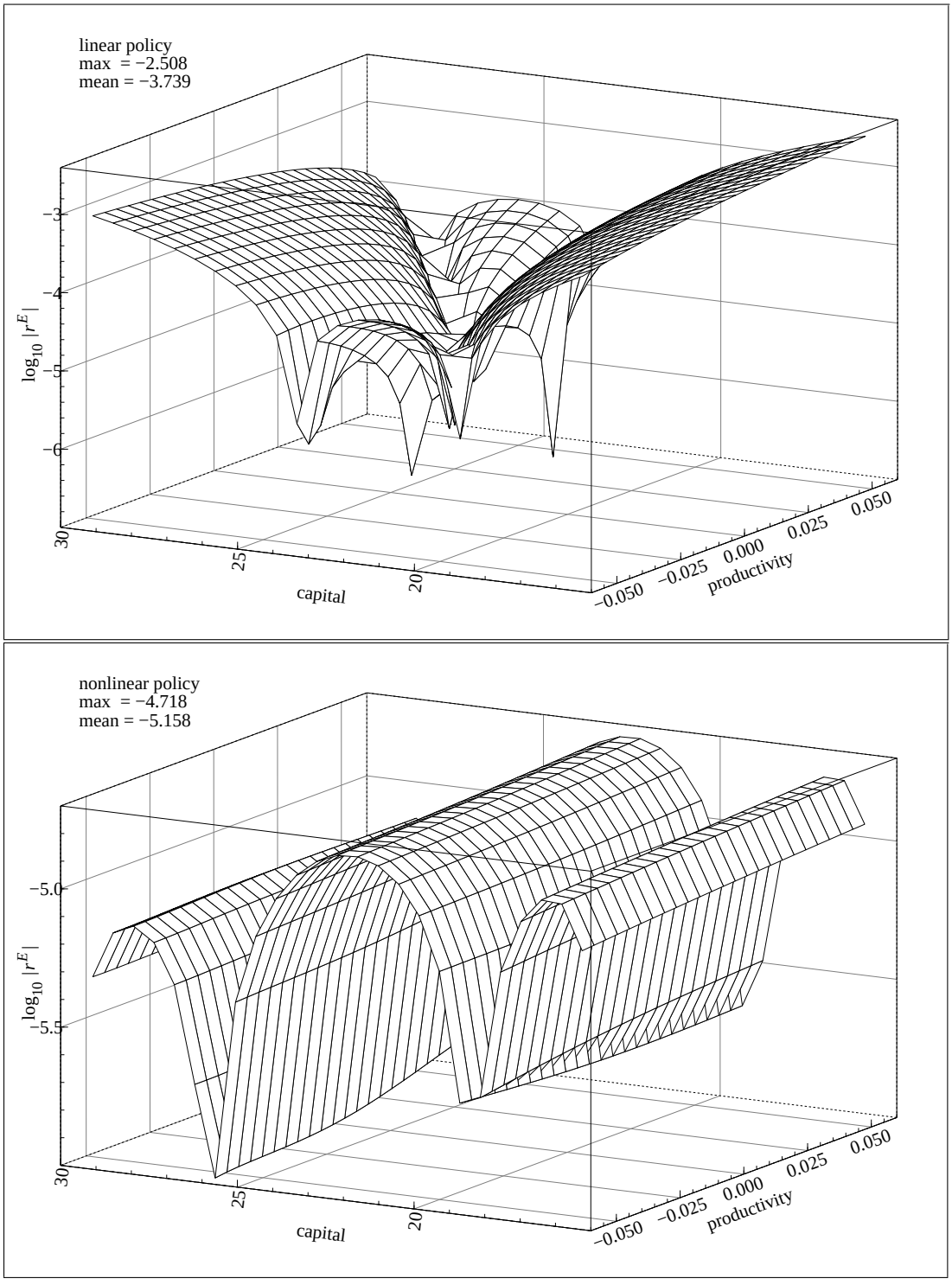


Figure 4.17: Policy Functions in the Risky Scenario

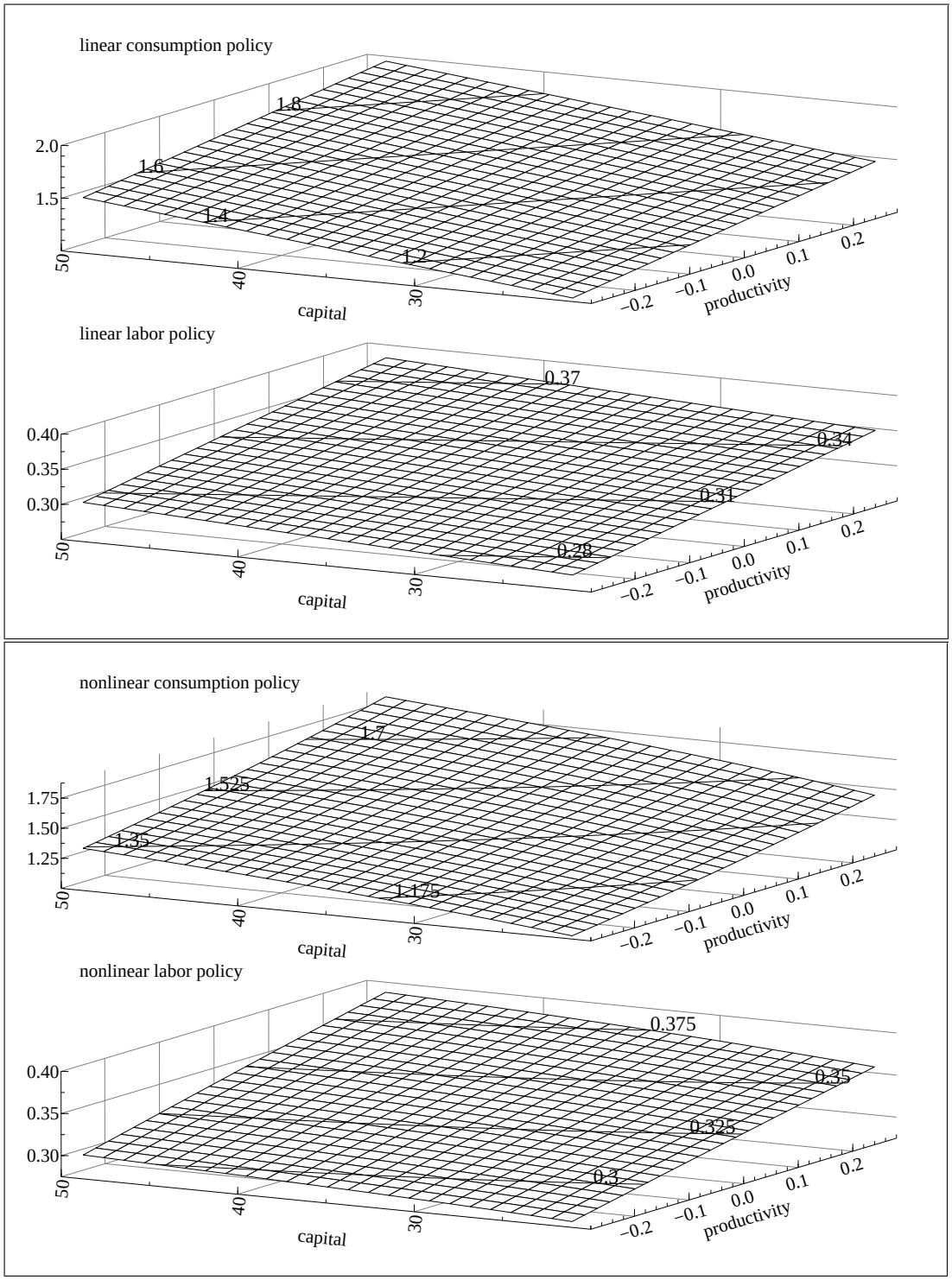


Figure 4.18: Euler Error in the Risky Scenario

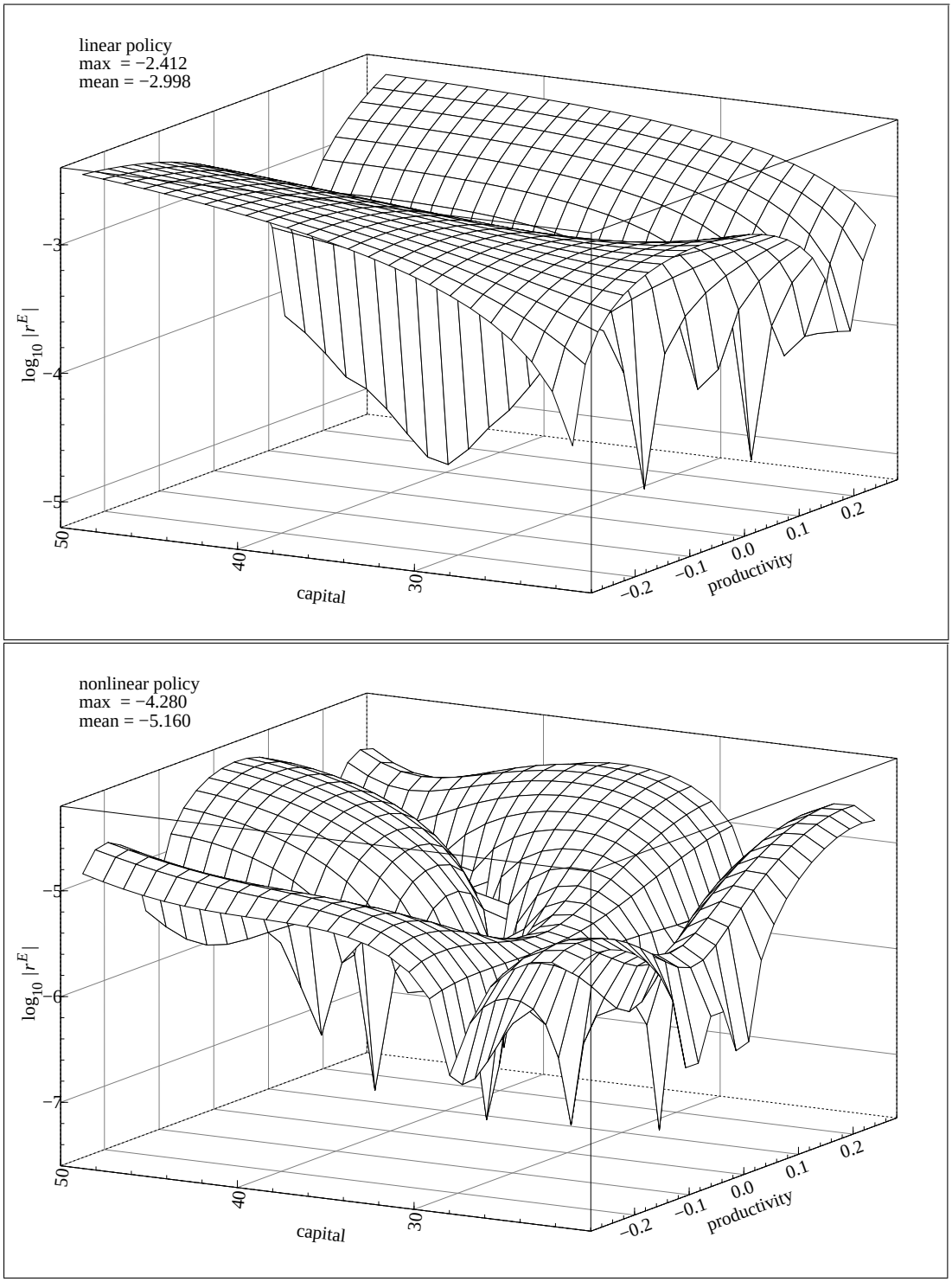


Figure 4.19: Smolyak Euler Error

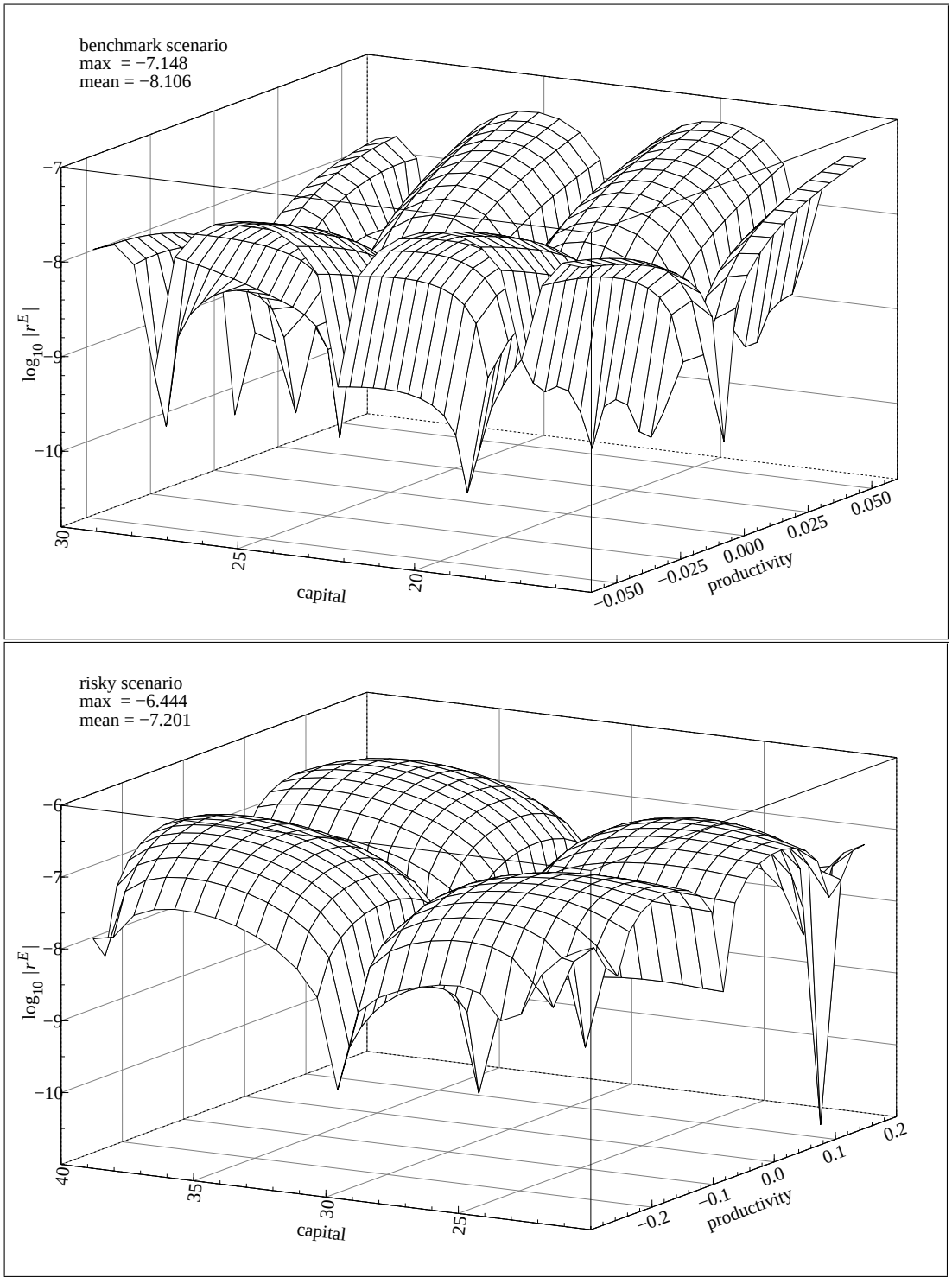
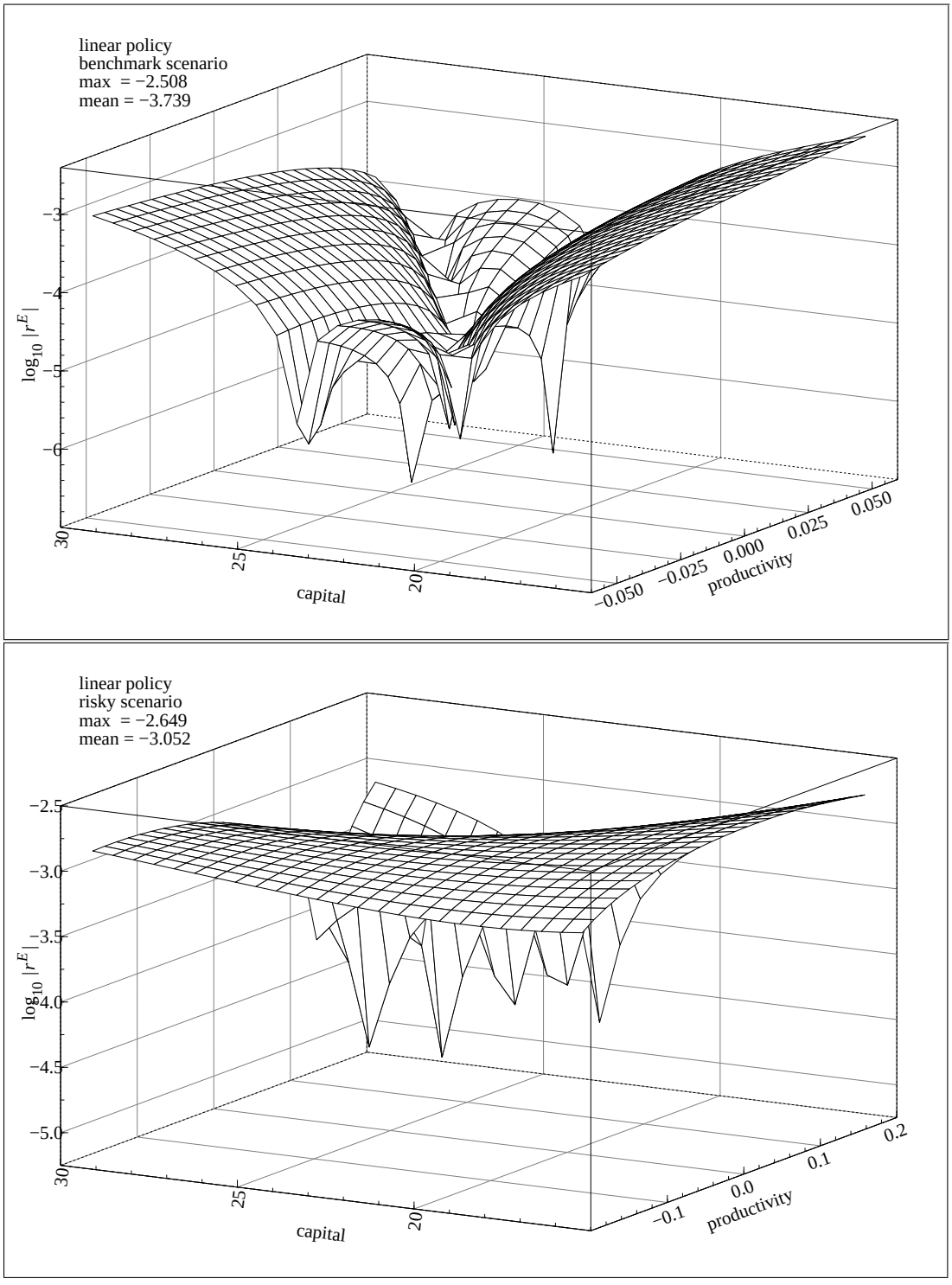


Figure 4.20: Euler Error of Linearization with Analytical Derivatives



parameter set. Whether these differences translate in a substantial estimation bias or not can hardly be judged by these plots alone. The first hint is to have a look on the implied Euler errors shown in figure 4.16. The linearized policy induces a maximal relative loss of around 1% of the consumption expenditures whereas the nonlinear policy maker suffers 100 times less from the use of the approximation instead of the exact optimal policy. The linearized solution deteriorates as expected when moving away from the steady state.
The solutions for the nonlinear scenario are given in figure 4.17. The absence of the certainty equivalence is represented in the Euler error of the linearized policy. They are larger around the steady state compared to the benchmark scenario.
The nonlinear policies were approximated by a 4^{th} degree Chebyshev polynomial on a 5×5 grid constructed by the Kronecker product. Figure 4.19 shows the Euler error for a 4^{th} level Smolyak approximation in both scenarios. In order to analyze the reduction of the computational effort by the Smolyak operator I calculated for the benchmark and risky scenario the nonlinear solutions with rising accuracy. The results are given in table 4.6. For the Smolyak operator

Table 4.6: Smolyak Reduction

Product Rule				Smolyak Operator			
Benchmark scenario							
Degree	Nodes	Max	Mean	Level	Nodes	Max	Mean
2	9	-3.81	-4.26	1	5	-3.22	-4.27
4	25	-5.51	-6.03	2	13	-5.13	-5.99
7	64	-7.96	-8.47	3	29	-7.15	-8.10
9	100	-9.47	-10.03	4	65	-9.97	-10.16
Risky scenario							
Degree	Nodes	Max	Mean	Level	Nodes	Max	Mean
1	4	-2.72	-3.07	1	5	-3.00	-4.28
3	16	-4.72	-5.64	2	13	-4.83	-5.86
5	36	-6.10	-7.12	3	29	-6.44	-7.20
8	81	-7.83	-8.70	4	65	-7.86	-8.35

the level q in equation (4.17) or (4.18) has to be chosen. The resulting number of nodes for the Kronecker product rule and the Smolyak operator are given in the column Nodes and the associated maximum and mean Euler error in the columns Max and Mean. In the benchmark parametrization the operator cuts the numbers of nodes by half to achieve approximately the same accuracy. In the risky scenario the dominance is lower and around 20%. The Smolyak reduction is lower in small dimensional problems and a higher reduction in percentages can be expected for larger models. The finite element approach of Fernádez-Villaverde and Rubio-Ramírez (2004b) needs 140 nodes for a maximal Euler error of around -5, whereas the Smolyak approach achieves this number with only 13 nodes.

Linearization is done by numerical first derivatives. I checked the linearization approximation with analytical derivatives with the *Mathematica* perturbation code from Aruoba, Fernádez-Villaverde, and Rubio-Ramírez (2003). For the benchmark case the numerical solution gives an accuracy of five digits for the policy functions and four digits for the implied transition matrix. For the risky case I get a policy function accuracy of 5 digits and 3 for the state transition. The implications for the Euler error can be seen in figure 4.20. In the benchmark scenario the use of the numerical derivatives has neither a visible nor measurable effect. In the risky scenario the numerical derivatives increase the maximal error from -2.649 to -2.412 and the mean error from -3.052 to -2.998.

4.7.2 Estimates

This subsection presents the estimation results. In the likelihood evaluation section the approximation quality of the filters is of central interest. Then the proposed Metropolis-Hastings algorithm is tested against the random walk variant in a Monte Carlo simulation study. Afterwards I present the structural parameter estimation with the genetic Metropolis-Hastings algorithm and the Gaussian filter. The last calculations compare the linearized and the nonlinear estimation according to their marginal likelihoods on the nonlinearly generated data set.

4.7.2.1 Likelihood

The likelihood values at the calibrated parameters in table 4.7 give the first idea of the approximation quality of the filters discussed in chapter 4.4. Both Gaussian

Table 4.7: Log Likelihood Values

Filter	Benchmark Scenario	Risky Scenario
Kalman filter	1,365.12	-150,575.32
Gaussian Kalman filter	1,368.23	1,209.91
Gaussian particle filter	1,368.19	1,211.06
bootstrap filter	1,369.92	1,215.68

filters are calculated with level 3 Smolyak quadrature for the time and the measurement update. The Gaussian particle filter uses 5,000 and the bootstrap filter 40,000 particles. The Kalman likelihoods are identical for the linearization with analytical and numerical derivatives. In the benchmark case the bootstrap filter gives virtually the same likelihood as the Gaussian filter. The importance sampling step in the Gaussian particle filter hardly improves the likelihood accuracy in the risky and not at all in the benchmark scenario. The Kalman likelihood is zero for the risky scenario whereas Fernádez-Villaverde and Rubio-Ramírez (2004a) report a log value above 1000. For the benchmark scenario Fernádez-Villaverde

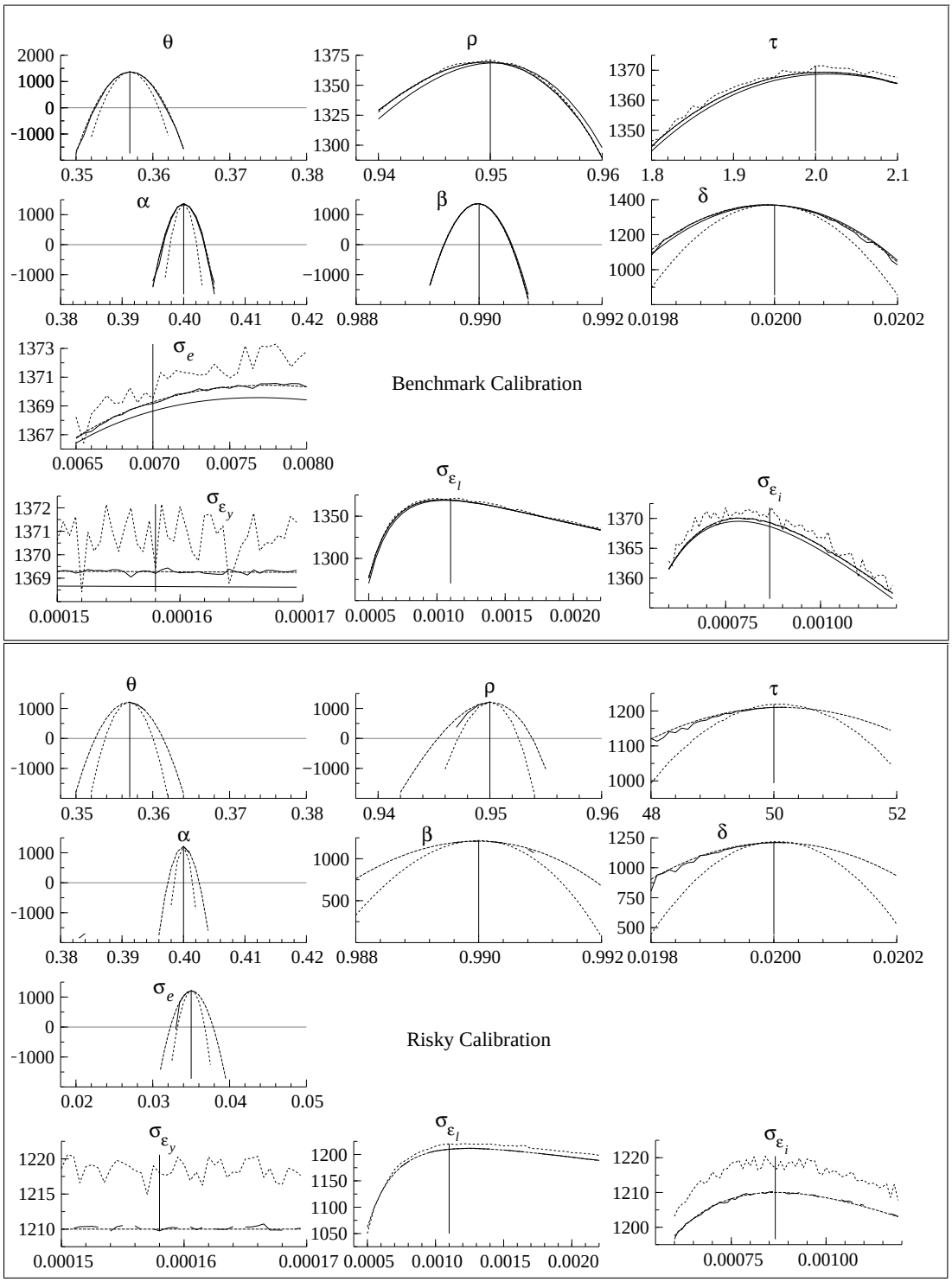
and Rubio-Ramírez (2004a) report 1462 and around 1000 for the bootstrap and Kalman likelihood, respectively.
Table 4.8 shows the convergence of the bootstrap filter when the number of particles is increased. The calculations are done with 50 replications. My results

Table 4.8: Convergence of the Bootstrap Filter

	Benchmark Scenario		Risky Scenario	
N	Mean	s.d.	Mean	s.d
10,000	1367.7	1.44	1213.5	9.9915
20,000	1368.0	0.74	1217.6	2.6814
30,000	1368.0	0.69	1218.0	1.6499
40,000	1368.1	0.62	1218.4	1.4650

show a lower standard deviation for the benchmark scenario whereas Fernádez-Villaverde and Rubio-Ramírez (2004b) report it the other way round.
The likelihood function is ten-dimensional and transversal cuts can be visualized by changing one parameter while keeping the other parameters at their calibrated values. Figure 4.21 shows likelihood values for the filters in the benchmark and the risky scenario. The dotted graphs belong to the Kalman filter, the varying graph to the particle filter. Both Gaussian filters produce virtually the same graphs without any detectable differences. The likelihood around the true autoregressive parameter ρ is rather flat and the estimation procedure will have problems to find the true value. It is also not informative for risk aversion parameter τ and shock variances σ_e, σ_{ϵ_l}, σ_{ϵ_i} and σ_{ϵ_y}. Parameters θ, α, β, δ show a strongly peaked likelihood and good estimates can be expected.
The likelihood values of the Gaussian filters are similar to the bootstrap filter values. For the Gaussian filter I used a third level Smolyak quadrature. The time update quadrature uses 39 nodes for the three dimensional integration over both states and the state shock. The measurement update needs 151 nodes for the five dimensional integrals of states and measurement shocks. The spectral interpolation of the policy function in the next period for each particle or node is the most demanding computation. The Gaussian filter is around 100 times faster then the bootstrap filter. This dramatic reduction is of course due to the reduction of 40,000 particles to 190 quadrature nodes.
The likelihood traversal cuts for the risky calibration are shown in the lower part of figure 4.21. The Kalman filter evaluations are not plotted since they completely miss the true parameters. The likelihood values of all three nonlinear filters are again very similar. Fernádez-Villaverde and Rubio-Ramírez (2004b) report values around 830 whereas I calculated much higher values. There are differences between the benchmark and the risky scenario concerning the flatness of the likelihood. In the risky scenario the output and investment measurement shock variances are clearly peaked, indicating a more informative likelihood. Good

Figure 4.21: Transversal Cuts Through the Likelihood



estimates can be expected for autoregressive parameter ρ, risk aversion τ and productivity shock σ_e.
Due to the similarity of the likelihood values for the Gaussian filter and the bootstrap filter I will only present estimates obtained by the simple Gaussian filter.

4.7.2.2 Normal Density

This section presents a Monte Carlo simulation study where the random walk Metropolis-Hastings (RWMH) algorithm is compared with the proposed genetic extension (GEMH).
The known target density is a ten dimensional multivariate normal distribution with the benchmark parameters as the expected values and the reported standard deviations of Fernádez-Villaverde and Rubio-Ramírez (2004b) as variances. The covariances are zero. Taking the standard deviations as variances is necessary because the squared estimated standard deviations result in a singular covariance matrix with a determinant of 3.8E-99.
There are two questions I investigate. How fast do the algorithms converge? And how good are the random numbers generated after convergence? For this purpose I ran 1,000 Metropolis-Hastings density estimations with the burn-in length detected automatically by a multiple reduction factor below 1.1. After the sequences have converged 50,000 subsequent draws are generated. The RWMH is run in three variants. The first (RWMHt) uses the scaled true target variance as the random walk variance. The other two (RWMHp) use the true variances multiplied with a small and a large random factor. The small factor is between 10^{-1} and 10^{1} and the large factor between 10^{-2} and 10^{2}. The GEMH is run with twenty parallel sequences. The innovation shock bound is $b = 10^{-5}$ and the parameter mixing factor is $\gamma^{GE} = 2.38/\sqrt{2 \times 10}$. Both number result in an acceptance ration of around 0.26. In a real application the true variances are not known and the RWMHt simulations are too optimistic and represent the optimal but not obtainable estimator quality. The perturbed variances correspond to the realistic situation of unknown variances.
The upper part of figure 4.22 plots the optimal RWMHt sequences. The solid straight lines represent the true parameter values and the dotted lines indicate a moving average. After some hundred draws the RWMHt sequence converge to the true values and after around 6,000 draws the diagnostic test detects the convergence of the density estimate. The right hand draws after the burn-in sequence can be taken as a sample from the target density. The lower part of figure 4.22 shows the RWMHp sequences with the little perturbed variances. The convergence towards the true values is as fast as for the RWMHt but the convergence of the density is detected only after 10,000 draws. Figure 4.23 shows one of the parallel sequences of the GEMH algorithm. Here the convergence towards the true value is faster if one counts only the draws of one sequence.

Figure 4.22: Random Walk Metropolis-Hastings

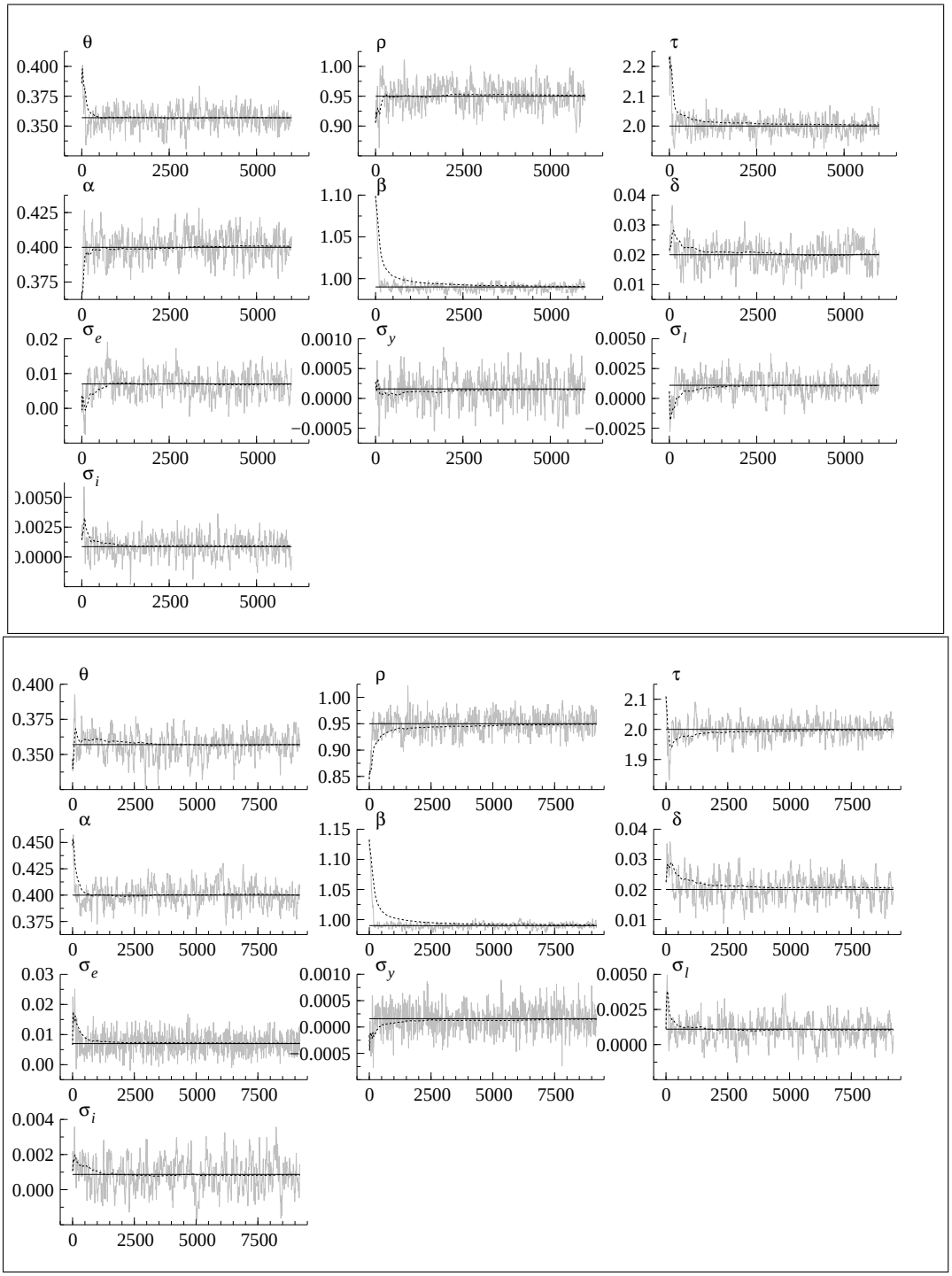
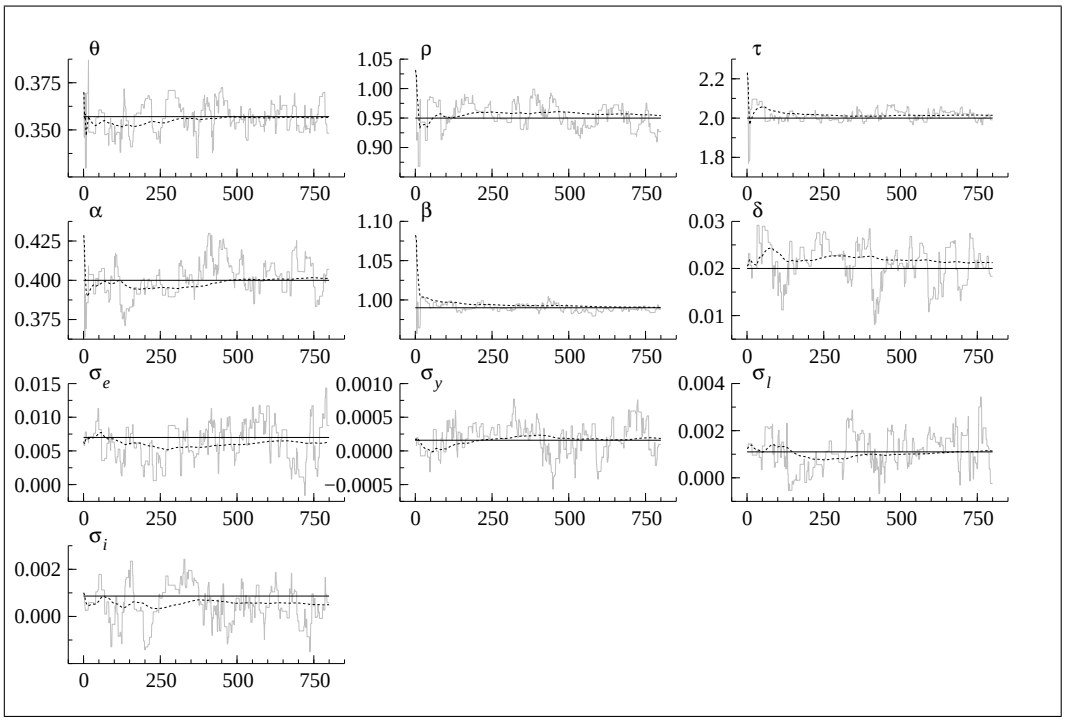


Figure 4.23: Genetic Metropolis-Hastings



This reflects the learning process effectuated by the mixing of parallel sequences as an addition source of innovation beside the pure random walk shocks. This learning procedure has to be done manually when several trial sequences are run sequentially and not simultaneously.

Table 4.9 shows the summary of the simulation results. In the upper part the length of the burn-in sequences are compared. It also shows two different simulations with regard to the dispersion of the start values around the true values. The burn-in length of the GEMH is the total number of draws in all parallel sequences.

The burn-in length of the RWMHp depends mainly on the start value dispersion and the variance perturbation does not influence it. Not surprisingly, the RWMHt shows the fastest convergence. The RWMHp algorithm with the little perturbed variances is comparable to the RWMHt performance. The burn-in length of the GEMH is independent of the start value dispersion.

Whether the GEMH is competitive with the RWMHp or not depends on ones view about a representative deviation of the tuned variances from the true ones. In my opinion it seems likely to use variances as far from the true ones as in the runs with a large factor perturbation. There is no sense in running trial sequences until the random shock variances represent the true ones since this estimate is the purpose of a final run. Or put it differently, the RWMHp sequences converge

Table 4.9: Metropolis-Hastings Performance

		RWMHp		RWMHt	GEMH
		small	large		
start value dispersion	Burn-in Lengths				
narrow	max	26,400	104,200	15,000	20,000
	top .95	11,800	27,200	8,800	16,000
	mean	6,944	13,651	5,215	13,720
	low .05	3,000	4,800	2,600	12,000
	min	1,200	2,400	800	8,000
wide	max	22,000	55,200	17,000	20,000
	top .95	12,400	28,600	10,400	16,000
	mean	7,332	14,423	6,343	14,144
	low .05	3,400	5,400	3,200	12,000
	min	1,400	2,600	1,600	8,000
Random Number Quality					
μ					
	mean bias	1.01	1.85	1.00	0.96
	variance	1.38	3.70	1.00	1.23
σ					
	mean bias	8.95	31.05	1.00	4.96
	variance	1.42	3.39	1.00	1.34

faster if the variances are tuned closed to the true ones in many trial runs or the sequences converge as slow as the GEMH due to wrong variances tuned in only a few runs.

If the performance of the highly perturbed RWMHp is taken as a realistic competitor for the GEMH then the result is that both algorithms achieve the same mean burn-in length. The difference between both algorithms is that in the trial sequences for the GEMH only two parameters have to be tuned. Moreover both parameters are the same for estimations independent of the size of the true variances. The RWMHp scale factor γ^{RW} needs some adjustment according to the size of the variance perturbation and the true variances.

The mean burn-in length is virtually the same for the RWMHp and GEMH. The variance of the GEMH burn-in length is much smaller and the RWMHp algorithm suffers from large outliers with the maximum as high as 104,200 draws. On the other side some estimations by the RWMHp algorithm converged much faster with a minimum burn-in length of 2,400. The result of this Monte Carlo simulation is that with respect to the burn-in length the GEMH is roughly comparable to the RWMHp. On average they converge after the same number of

draws but the GEMH does not need trial sequences to find suitable random walk shock variances.
The lower part of table 4.9 shows the random number quality. The mean of 50,000 draws after the burn-in sequences is the estimated expected value. For each of 1,000 simulations I calculated the absolute deviation of this estimate from the true value. The mean of these deviations is the mean bias which together with the variance characterizes the estimation quality. These measures are calculated for the expected values and the variances of the multinormal distribution. Then I divide these numbers by the corresponding optimal performance of the RWMHt simulations and take the mean over all 10 parameters.
The mean estimates of the small perturbed RWMHp and GEMH are comparable with the RWMHt estimates. The highly perturbed RWMHp shows the weakest performance. This also holds for the variance of the mean estimates. Estimates of the variances are not as good as the mean estimates for all algorithms but GEMH again performs better than RWMHp.

4.7.2.3 Posterior

This section reports the estimates of the posterior density of structural parameters of the model at hand from a sample with 100 observations. Two data sets are simulated - one with the benchmark and one with the risky parameter values. The parameters are calibrated for quarterly data and a sample represents 25 years. Both data sets are generated with the nonlinear solution. All posterior estimates are obtained by the Gaussian filter as the nonlinear estimator. As the linear estimator I take a linear perturbation with numerical derivatives and the Kalman filter.
The convergence test of the genetic Metropolis-Hastings algorithm consists of the acceptance ratios, the multivariate scale reduction factor R and the autocorrelation functions. The acceptance ratios have to be around .3 for each parallel sequence. The autocorrelation functions for each sequence have to fall fast to around .5 within 40 lags. They are very similar across sequences and parameters and I report only one representative autocorrelation function. The scale reduction factor R must fall below 1.1. The converged parallel sequences are stacked and a histogram represents the density estimate of the posterior.
An important decision is the choice of the start values for the sequences. Two different steps are necessary to obtain a posterior estimate. The first step has to find the modes and the second step samples around them to generate representative draws. For the maximization step the optimal mixing parameter and the shock variances are larger than for the sampling draws. I used the resampling step of the bootstrap filter for a posterior based mixing of the parallel parameter vectors. It accelerates the maximization substantially. The maximization procedure finds the mode within some thousand likelihood evaluations. The mode seems to be

Table 4.10: Parameter Estimates

	True	mean		std.dev.	
		Nonlinear	Linear	Nonlinear	Linear
	Benchmark Scenario				
θ	0.35	0.358757	0.358176	1.83E-03	1.63E-03
ρ	0.95	0.921978	0.931711	1.90E-02	1.35E-02
τ	2.00	2.873962	2.637364	4.89E-01	3.75E-01
α	0.40	0.404878	0.403540	4.75E-03	4.20E-03
β	0.99	0.989154	0.989405	8.90E-04	7.71E-04
δ	0.02	0.021091	0.020784	1.06E-03	9.25E-04
σ_a	0.007	0.006349	0.006242	4.71E-04	3.92E-04
σ_y	0.000158	0.000672	0.000719	3.58E-04	3.34E-04
σ_l	0.0011	0.001271	0.001272	9.00E-05	8.00E-05
σ_i	0.000866	0.000592	0.000570	2.57E-04	2.50E-04
	Risky Scenario				
θ	0.35	0.356555	0.359832	4.47E-04	3.56E-03
ρ	0.95	0.950678	0.942740	8.57E-04	6.74E-03
τ	50.00	49.385802	62.910866	9.01E-01	2.07E+01
α	0.40	0.399076	0.408750	1.25E-03	9.20E-03
β	0.99	0.990152	0.988453	2.09E-04	1.79E-03
δ	0.02	0.019809	0.025163	2.62E-04	2.26E-03
σ_a	0.035	0.035053	0.031274	1.17E-04	2.11E-03
σ_y	0.000158	0.000160	0.000674	6.10E-06	4.58E-04
σ_l	0.0011	0.001080	0.001351	2.55E-05	9.24E-05
σ_i	0.000866	0.000877	0.000994	1.89E-05	2.15E-04

unique for the simple model at hand. For multimodal densities specialized mixing strategies are available, see ter Braak (2004).

I do not present the maximization sequences but only the sampling step with start values narrowly dispersed around the true parameters. Starting from any parameter vector needs only more draws in addition to the burn-in draws thereafter without changing the density estimate. However, generating posterior histograms needs many more draws than finding the mode of the density.

Figure 4.24 present the diagnostics for the benchmark and the risky scenario estimation. The histograms are drawn for the obtained posterior values after the burn-in sequences. The vertical black lines indicate the last values of the parallel sequences. For both data sets the sampling takes place within some units of the log posterior. The acceptance ratio is around .3 and the representative autocorrelation function is falling fast. The scale reduction factor R is below 1.1 and indicates convergence of the density estimates.

Figure 4.24: Convergence Diagnostics

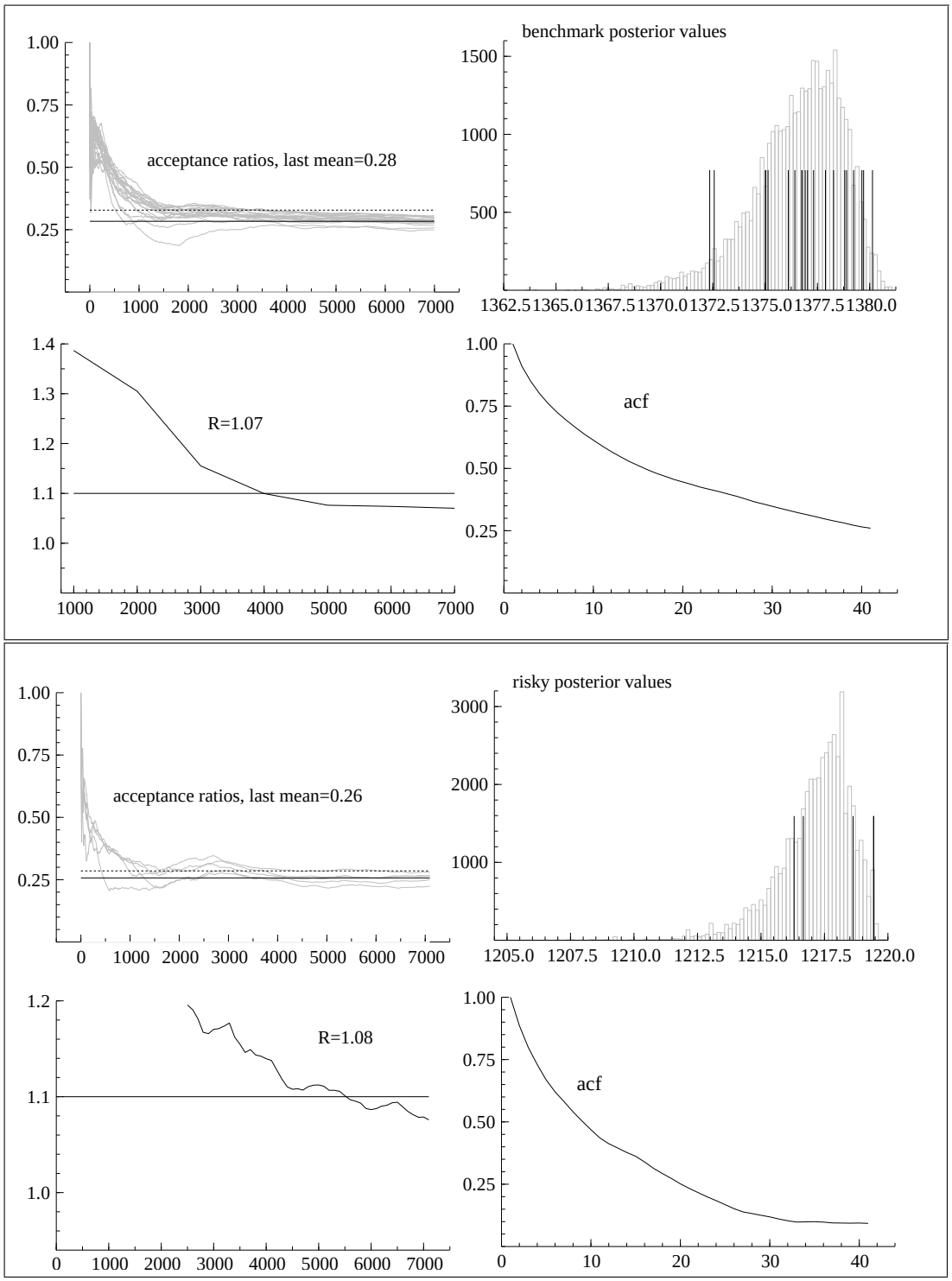


Figure 4.25: Parallel Metropolis-Hastings Burn-In Sequences

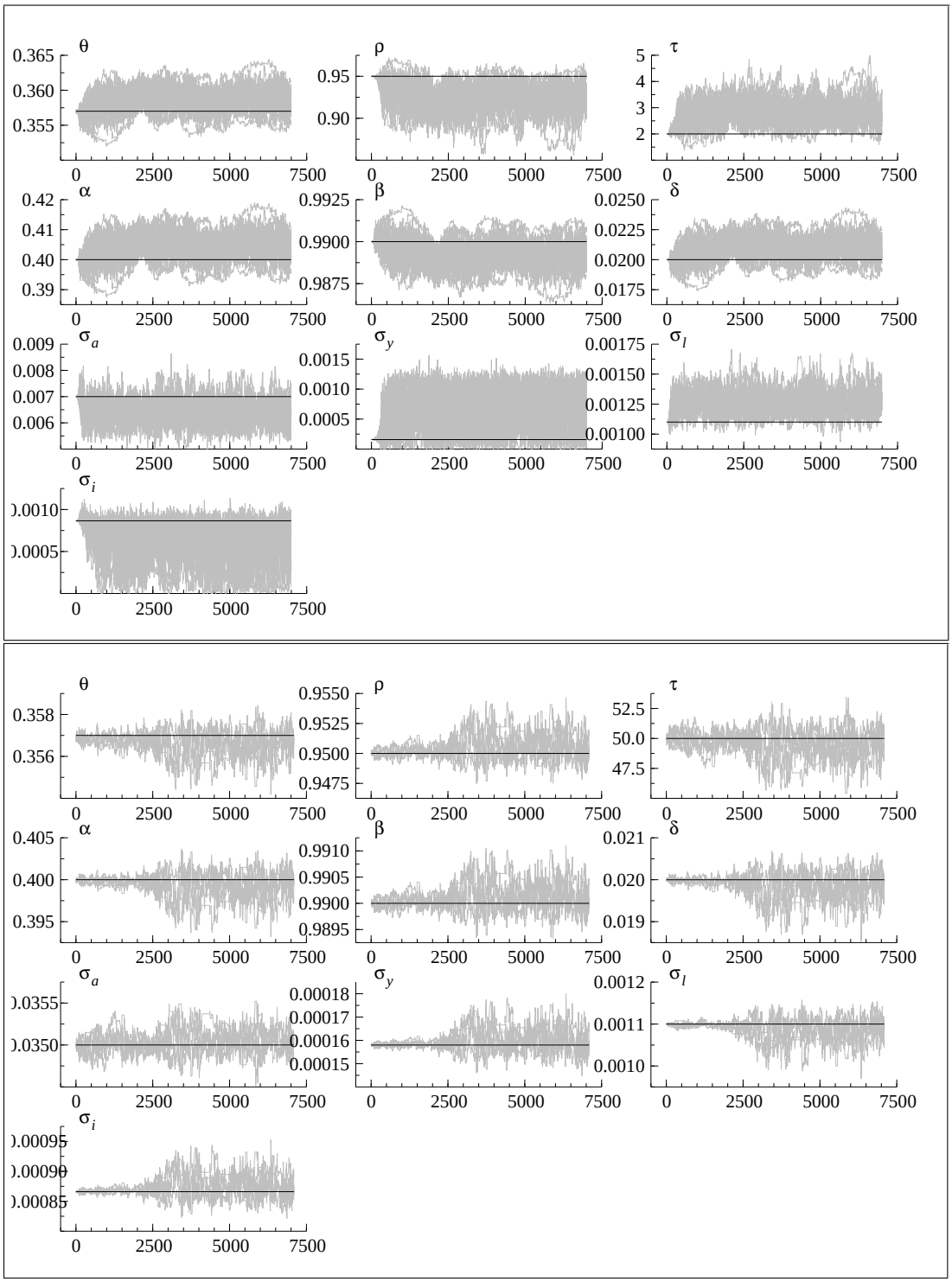
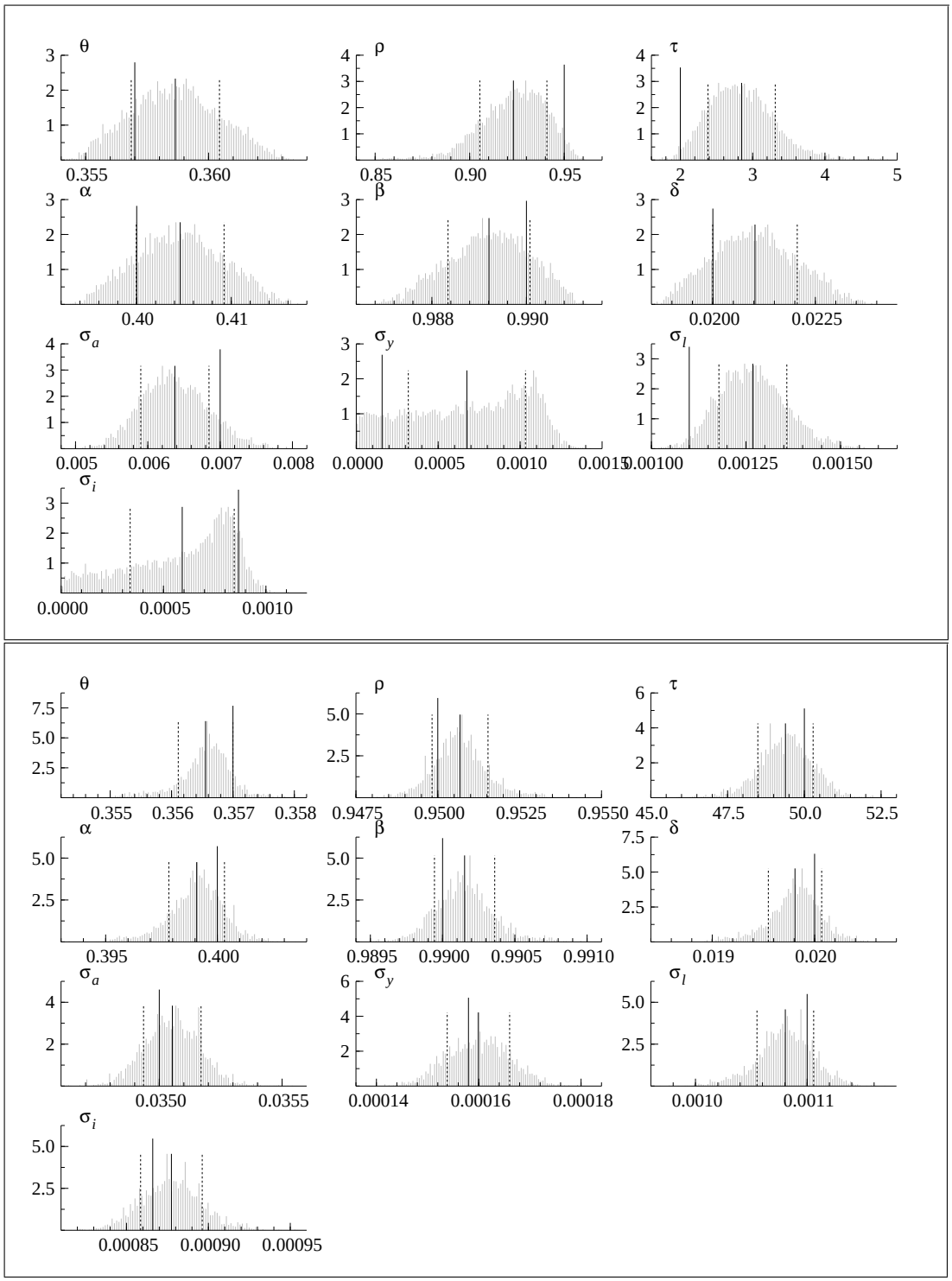


Figure 4.26: Estimated Posterior Densities



The optimal parameters γ^{GE} and b were found to be independent of the estimator since the ones tuned in the linear estimation of the benchmark scenario were also optimal for the nonlinear estimations. I used the optimal mixing factor $\gamma^{GE} = 2.38/\sqrt{2 \times 10}$ scaled by 0.6 and a uniform shock bounds of $b = 10^{-9}$. The linear tuning runs took only some minutes and simplified the estimation process substantially since expensive nonlinear runs were superfluous. Whether this independence of the algorithm parameters from the estimator carries over to other models is to be tested. However, there is not much scope for tuning the parameters since γ^{GE} has an strong effect on the acceptance ratio. Changing the optimal mixing factor by a factor between 0.2 and 1 changed the acceptance ratio dramatically. The variance parameter of the random shocks b is of minor importance without to much influence on the acceptance ratio since most of the innovation comes from the mixing of the parallel sequences. In fact the number of parallel sequences is another but apparently less important parameter. It should depend on the number of parameters to be estimated and the number of modes of their posterior density. I have run the model estimations with 20 and 5 parallel sequences for the different parameter sets without noticeable implications.
In the benchmark case the potential scale reduction factor R fell below 1.1 after 5,000 draws from 20 sequences and in the risky case 7,000 draws from 5 sequences were needed. Afterwards the sequences were run until 50,000 posterior draws were generated.
Figure 4.25 shows all parallel burn-in sequences for the nonlinear estimations. The black horizontal lines represent the true parameters. None of the sequences is trending anymore and taken each separately they look similar to the middle graph in figure 4.14. The linear Kalman filter sequences are not shown. They move away from the true parameters towards their biased estimates.
Figure 4.26 shows the nonlinearly estimated parameter posteriors. The long black vertical line indicates the true values, the shorter lines the mean and the dotted lines bound two standard deviations around the mean. Table 4.10 presents the first moments of the posterior as point estimates. The uncertainty of these estimates is given by the second moment of the posterior. The Kalman filter estimates of the linearized model are listed as a comparison.
The benchmark parameter estimates are similar for the linear and nonlinear estimation. The standard deviations of the linear estimates are slightly smaller but the not maximum of the likelihood. The estimates of risk parameter τ, autocorrelation ρ and the shock variances are biased and unprecise. This is different to Fernádez-Villaverde and Rubio-Ramírez (2004b) who report better nonlinear estimates. This may be driven by the sloppy solution of the singularity problem by measurement errors, a bug in the code or the simplicity of the Gaussian filter. A bug in the code and the Gaussian filter are unlikely the cause of the bias since the risky parameter point estimates are good and in line with Fernádez-Villaverde and Rubio-Ramírez (2004b). Moreover the Gaussian filter gives at least at the true benchmark parameters the same likelihood value as the bootstrap filter.

Further experience with the code, other models and filters should resolve this puzzle.
The risky parameter point estimates are close to the true values and they are very precise with all true values lying within the standard deviation bounds. Here the nonlinear estimation clearly outperforms the linear one. The linear estimation is biased in τ, δ and all shock standard deviations. Fernádez-Villaverde and Rubio-Ramírez (2004b) report similar point estimates but around 10 times smaller standard deviations. Some differences might of course be due to different simulated data sets.
Nonlinear and linear estimates imply different moments of observables. Fernádez-Villaverde and Rubio-Ramírez (2004a) simulate the model and find this difference to be substantial. This is critical since correlations among observables are often used as an informal model selection criterion.

4.7.2.4 Marginal Likelihood

Table 4.11: Marginal Likelihood

	Benchmark Scenario			Risky Scenario		
	@true	1368.2	1365.1		1209.9	-143,651.0
	max	1374.4	1373.7		1213.2	1190.3
p	Δ	Nonlinear	Linear	Δ	Nonlinear	Linear
.1	3.4	-71.5	-74.8	28.8	-111.6	-140.4
.2	3.4	-70.7	-74.1	28.8	-110.9	-139.7
.3	3.5	-70.2	-73.7	28.8	-110.5	-139.3
.4	3.7	-69.6	-73.4	28.8	-110.2	-139.0
.5	3.8	-69.3	-73.1	28.8	-110.0	-138.8
.6	4.0	-68.9	-72.9	29.0	-109.6	-138.6
.7	4.2	-68.5	-72.7	29.0	-109.4	-138.5
.8	4.3	-68.0	-72.3	29.1	-109.3	-138.3
.9	4.8	-67.0	-71.8	29.2	-109.0	-138.2

The econometric variable of interest in this paper is the marginal likelihood given in table 4.11. It shows the linear versus nonlinear model selection for the benchmark data set and the risky one. In the first two rows there are the log likelihoods at the true parameters and the log likelihood maxima for all sequences. In the lower part there are the marginal log likelihoods for different quantiles. The nonlinear likelihood at the true values is above zero, higher than the Kalman likelihood and higher at maximum for both data sets. The nonlinear marginal likelihood is higher for the benchmark data set and much higher for the risky set. The nonlinear model is detected to fit both data sets better than the lin-

earized model even if the parameters imply a rather linear policy function in the benchmark case.

4.8 Software

The results were generated using the matrix language Ox version 3.40, see Doornik (2002).[4] It is faster than Matlab and has a C style syntax which simplifies portations to C. The hardware I used is a Pentium 4, 3GHz.
The code consists of three pieces. The first part is a general function approximation tool with 500 lines. It fits Smolyak Chebyshev polynomials to function evaluations on a sparse grid. The second part with 1,200 lines is an integration toolbox which delivers a grid of nodes and associated weights for different densities and approximation levels. The nodes and weights can be uniform, normal, lognormal, t-student, gamma or beta distributed. It delivers Kronecker, nonnested Smolyak nodes and weights as well as nested hierarchical nodes and weights according to Genz and Keister (1996b). Any other weight function for which a univariate operator exists or is programmable can be plugged in. Other than normally distributed shocks can be of use for financial market applications or specification tests. The last and main part of the code with 3,500 lines contains the linear and nonlinear solution routines, various filters, the genetic Metropolis-Hastings algorithm, report functions and the marginal likelihood integration. A model file defines first order conditions $\boldsymbol{f}$, state transition functions $\boldsymbol{g}$, measurement functions $\boldsymbol{m}$, the Euler error and various options for the algorithms.
The running times including burn-in sequences for the linear estimation are 10 min and 2 h 20 min for the benchmark and risky parametrization. The nonlinear estimation took 10 and 15 h for $5,000 \times 20 + 50,000$ and $7,000 \times 5 + 50,000$ likelihood evaluations. The tuning sequences for the genetic Metropolis-Hastings parameters took some minutes with the linear Kalman filter.
Smolyak integration can be expected to work efficiently for at least 30 dimensions. Petras (1999) approximated a present value integral with 360 dimensions of different importance in 2 minutes on a 400MHz Pentium machine. Krüger and Kübler (2004) report 14 h for one Smolyak based solution of an OLG model with 30 generations. They used the closed form solution formulas of the Chebyshev coefficients on a Pentium 4, 3GHz. The switch to a matrix based coefficient calculation accelerated in my code the approximation by a factor between 10 and 100.
The interpolation step for the Smolyak Chebyshev polynomials is coded in C. It is the main bottleneck of the whole estimation process and I obtained in C a poor two fold speed up compared to the Ox code for low dimensions and a bit more

[4] Matlab is a trademark of the The MathWorks, Inc. Mathematica is produced by Wolfram-Reasearch, Inc. Maple is a trademark of Waterloo Maple Inc.

for higher dimensions. A specialized routine is the first candidate for a further acceleration.
This is especially urgent for the bootstrap filter. The estimation for the simplest model with the nonlinear bootstrap filter takes 20 seconds for each likelihood evaluation. This sums up for 5 sequences to $(7,000 \times 5 + 50,000) \times 20$ sec ≈ 472 h or 20 days. The Fortran finite element bootstrap filter in Fernádez-Villaverde and Rubio-Ramírez (2004b) takes 88 h for 50,000 draws on a Pentium 4, 3 GHz.
I wrote also the resampling step for the bootstrap filter in C since it needs some nested loops and a vectorization is not possible. The speed gain for each period likelihood contribution is substantial and 40,000 particles are resampled in 0.016 seconds compared to 3.1 in the original Matlab version. The underlying Matlab function `residualR.m` is available on Nando de Freitas' homepage and is a joint work with Arnaud Doucet accompanying the paper van der Merwe, de Freitas, Doucet, and Wan (2001).
I translated the `spgetseq.m` Matlab function by Klimke (2004) for the generation of the Smolyak indices $\mathbf{i}$ in equation (4.18). It is a part of the Smolyak based linear spline interpolation toolbox. This toolbox cannot extrapolate and is therefore not suited for solving economic models. Extrapolation is critical but arises when the states at the borders of the approximation space transit to the next period states outside the border. It sometimes caused the solution algorithm to diverge in the beginning of the maximization draws of the Metropolis-Hastings sequences. This in turn is not critical and after a higher likelihood value is reached divergence does not occur anymore.
I extracted the nodes and weights generation form the fully symmetric Gaussian quadrature code written by Genz and Keister (1996b). Their Fortran code is available at Alan Genz's homepage. A Smolyak construction of nonnested univariate quadrature nodes and weights works equally well for nonadaptive integration. This is done with code based on the functions `qnwnorm.m`, ..., `qnwbeta.m` of the *Compecon* Matlab toolbox accompanying Miranda and Fackler (2002) which at the core are translations of quadrature routines in Press, Flannery, Teukolsky, and Vetterling (1988). I also extended the *Compecon* Matlab routines to generate Smolyak based nodes and weights for the paper of Heiss and Winschel (2005).
The linear perturbation solution needs an ordered generalized Schur decomposition. Ox does not provide an ordering option for its Schur decomposition in `decschurgen` and I translated the `qzswitch.m` Matlab function of Christopher Sims.
The marginal likelihood calculation is based on the Matlab function `marginal.m` of Juan F. Rubio-Ramírez. It is part of a Bayesian econometrics course downloadable from his homepage.

4.9 Conclusion

The framework allows a very general structural analysis of dynamic nonlinear econometric models. Larger models than the one used here might take some days of estimation on a stand-alone computer. Some few lines of code to exchange the current draws of the genetic Metropolis-Hastings algorithm allow the estimation to be done on a parallel computer cluster. Compared to the approach in Fernádez-Villaverde and Rubio-Ramírez (2004b) my code is around 20 times faster. Since it is written in an interpreted language a further 10 fold speed up in a C implementation should be possible. The speed gain of my approach compared to Fernádez-Villaverde and Rubio-Ramírez (2004b) can therefore be expected to be around 200.
The approximation quality of the model solution with Chebyshev polynomials is good and the Smolyak operator substantially reduces computations for approximation and integration already for the smallest possible model.
The nonlinear estimates of the risky parameters are good and better than the benchmark estimates. The estimates of the benchmark parameters are biased and unprecise for the risk parameter τ, autocorrelation ρ and shock variances.
The Gaussian filter is fast and a Gaussian sum approximation of the posterior state densities is a possible cheap road towards the accuracy of the bootstrap filter. Further experience with both filters is desirable for other models.
The genetic extension of the Metropolis-Hastings algorithm is a useful alternative to the random walk algorithm with one sequence and pure random walk shocks. Within the estimation process its handling is rather uncomplicated and the parameters for the algorithm can be tuned fast by a linear estimation. The genetic extension allows an integrated global hill climbing as a first step of a density estimation before sampling around the mode sets in.
The model selection criterion can be calculated fast and it detects the nonlinear nature of the simulated data even if the policy function is almost linear.
The structure of the statistical decision theory and its variable of interest can be fruitfully applied to econometrics for the calculation of specification tests and out-of-sample forecasts.

Chapter 5

Smolyak Cubature for Multiple Integration in Estimation Problems

5.1 Introduction

Many econometric models imply likelihood and moment functions that involve multidimensional integrals without analytically tractable solutions. This problem arises frequently in microeconometric models in which all or some of the endogenous variables are only partially observed. Other sources include unobserved heterogeneity in nonlinear models and expectations of agents.
There are different approaches for numerical integration. It is well known that Gaussian quadrature performs very well in the case of one-dimensional integrals of smooth functions (Butler and Moffit 1982). Quadrature can be extended to multiple dimensions and in this case it is also called cubature. The most direct extension is a tensor product of one-dimensional quadrature rules. However, computing costs rise exponentially with the number of dimensions and become prohibitive for more than four or five dimensions. This phenomenon is also known as the curse of dimensionality.
The computational challenges of high dimensional integrands led to the advancement and predominant use of simulation techniques for the numerical approximation of multidimensional integrals in the econometric literature, see for example McFadden (1989) or Börsch-Supan and Hajivassiliou (1993). Hajivassiliou and Ruud (1994) provide an overview over the general approaches of simulation and Train (2003) provides a textbook treatment with a focus on discrete choice models, one of the major classes of models for which these methods were developed and frequently used.
This chapter is based on joint work with Florian Heiss and proposes and investigates the performance of a different approach that can be traced back to Smolyak (1963b). It has been advanced in recent research in numerical mathematics, see for example Novak and Ritter (1999). It is based on one-dimensional Gaussian

quadrature but extends it to higher dimensions in a more careful way than the tensor product rule. This dramatically decreases computational costs in higher dimensions.
Just like Gaussian cubature and simulation, Smolyak cubature evaluates the integrand at certain points and calculates a weighted average of these function values. The difference is how these points and weights are derived. The Smolyak approach is a general method for multivariate extensions of univariate operators and is applicable not only to integration. Another example is function approximation which was used for the solution of an overlapping generations model by Krüger and Kübler (2004).
After introducing the Smolyak approach, the results of Monte Carlo experiments are presented. We directly address the question of interest for estimation: Which method delivers the best estimates with a given amount of computing costs? The experiments are based on a panel data random parameters logit models which are widely used in the applied discrete choice analysis. We vary the panel data dimensions, the number of alternatives, the dimension of unobserved taste components and the parametrization of the data generating process. The results show that the Smolyak-based cubature methods clearly outperform simulations based on both random number generators and the modified Latin hypercube sampling proposed by Hess, Train, and Polak (2005).
This chapter is structured as follows: Section 5.2 briefly discusses the circumstances in which multiple integrals evolve in estimation problems. It then introduces an example, the random parameters logit model, in somewhat more detail since it will be used in the Monte Carlo experiments. Section 5.3 discusses the general approaches to numerical integration and introduces Smolyak-based cubature. Section 5.4 presents the Monte Carlo design and results. Section 5.5 concludes.

5.2 Econometric Models Requiring Numerical Integration

The log-likelihood function of microeconometric models can typically be written as a sum over a number N of independent log-likelihood contributions $\log(\ell_i(\boldsymbol{\theta}; \text{data})$. Maximum likelihood defines the estimated parameter vector $\hat{\boldsymbol{\theta}}$ as

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{N} \log(\ell_i(\boldsymbol{\theta})). \tag{5.1}$$

The discussion is focused on maximum likelihood estimation, but the same problems and approaches are applicable for other methods like GMM or Bayesian analysis. In many models the likelihood contributions $\ell_i(\boldsymbol{\theta})$ involve multiple integrals which cannot be expressed in closed form and must be evaluated numerically.

The approximation algorithm used for this numerical evaluation is essential for the estimation task. Numerical maximization involves repeated evaluations of the likelihood function. Each evaluation in turn involves solving N multiple integrals, where N in practice can be several thousand. While the ongoing increase in computational power makes widespread use of such models feasible, the computational costs are still high and sometimes prohibitive. Instead of compromising on model specification or approximation quality, it is therefore important to choose an efficient method of numerical integration for a given model and accuracy.
There are various reasons why microeconometric models imply multiple integrals in $\ell_i(\boldsymbol{\theta})$. Typically, they represent the expectation of a function over several random variables. A major reason for their presence is that many models are specified in terms of latent random variables for which the observed endogenous variables provide only a partial indication. Another source in nonlinear models is an error term with a mixture distribution such as random effects or error components models which can be estimated by calculation of integrated likelihood functions. Finally, dynamic optimization models naturally involve multiple integrals, see for example Eckstein and Wolpin (1999). For a more general presentation see for example Hajivassiliou and Ruud (1994) or Gouriéroux and Monfort (1996), both from a perspective of simulation.
The random parameters logit (RPL) or mixed logit model is widely used for studying choices between a finite set of alternatives. See McFadden and Train (2000) for an introduction to this model and a discussion of its estimation by simulation methods. Suppose discrete choices of N individuals are observed. The data has a panel structure, so that each of the subjects makes T choices. In each of these choice situations, the individual is confronted with a set of J alternatives and chooses one of them. These alternatives are described by K exogenous attributes. The $(K \times 1)$ vectors $\mathbf{x}_{itj}$ collect these attributes of alternative $j = 1, ..., J$ in choice situation $t = 1, ..., T$ of individual $i = 1, ..., N$.
Random utility maximization (RUM) models of discrete choices assume that the individuals make their choices by evaluating the utility that each of the alternatives yields and then picking the one with the highest value. The researcher obviously does not observe these utility levels. They are modelled as latent variables for which the observed choices provide an indication. Let the utility that individual i attaches to alternative j in choice situation t be represented by the random coefficients specification

$$U_{itj} \quad = \quad \mathbf{x}'_{itj}\boldsymbol{\beta}_i + e_{itj}. \tag{5.2}$$

It is given by a linear combination of the attributes of the alternative, weighted with individual-specific taste levels $\boldsymbol{\beta}_i$. These individual taste levels are distributed across the population according to a parametric joint p.d.f. $f(\boldsymbol{\beta}_i; \boldsymbol{\theta})$ with support $\Psi \subseteq \mathbb{R}^K$. The i.i.d. random variables e_{itj} capture unobserved utility components. They are assumed to follow an Extreme Value Type I (or Gumbel)

distribution. Note that this model can be generalized, for example, the distribution $f(\boldsymbol{\beta}_i;\boldsymbol{\theta})$ can be specified as a function of observed individual characteristics. Our goal is to estimate the parameters $\boldsymbol{\theta}$. Let y_{itj} denote an indicator variable that has the value 1 if individual i chooses alternative j in choice situation t and 0 otherwise. Denote the vector of observed individual outcomes as $\mathbf{y}_i = [y_{itj}; t = 1, ..., T, j = 1..., J]$ and the matrix of all exogenous variables as $\mathbf{x}_i = [\mathbf{x}_{itj}; t = 1, ..., T, j = 1..., J]$. Then, the probability that the underlying random variable $\mathbf{Y}_i$ equals the observed realization $\mathbf{y}_i$ conditional on $\mathbf{x}_i$ and the taste levels $\boldsymbol{\beta}_i$ can be expressed as

$$P_i^*(\boldsymbol{\beta}_i) = \Pr(\mathbf{Y}_i = \mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\beta}_i) \quad = \quad \prod_{t=1}^{T} \frac{\prod_{j=1}^{J} \exp(y_{itj}\mathbf{x}'_{itj}\boldsymbol{\beta}_i)}{\sum_{j=1}^{J} \exp(\mathbf{x}'_{itj}\boldsymbol{\beta}_i)}. \tag{5.3}$$

Suppose the regularity conditions given by McFadden and Train (2000) hold. The likelihood contribution of individual i as a function of $\boldsymbol{\theta}$ can be written as

$$\ell_i(\boldsymbol{\theta}; \mathbf{Y}_i) = \Pr(\mathbf{Y}_i = \mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\theta}) \quad = \quad \int_{\Psi} P_i^*(\boldsymbol{\beta}_i) f(\boldsymbol{\beta}_i;\boldsymbol{\theta})\, d\boldsymbol{\beta}_i. \tag{5.4}$$

A solution for this K-dimensional integral does not exist in closed form and has to be approximated numerically.

5.3 Numerical Integration in Multiple Dimensions

There are several methods to numerically approximate an integral of a function g over a D-dimensional vector $\mathbf{z}$.

$$I^D[g] = \int_{\Omega} g(\mathbf{z})\, w(\mathbf{z})\, d\mathbf{z}, \tag{5.5}$$

where $w(\mathbf{z})$ is some weighting function. As discussed above, in estimation problems, the integral often represents an expected value of g so that $w(\mathbf{z})$ is a p.d.f. and Ω its support. A computationally feasible approach that is common to all methods discussed in this chapter is to approximate the integral as a weighted sum of R integrand evaluations at certain points, referred to as nodes:

$$I^D[g] \approx \sum_{r=1}^{R} g(\mathbf{z}_r) w_r, \tag{5.6}$$

where w_r is the weights of node $\mathbf{z}_r$. The methods differ in the way they derive the nodes $\mathbf{z}_r$ and weights w_r. For Monte Carlo integration, the $\mathbf{z}_r$ are equally weighted draws from the density $w(\mathbf{z})$ with $w_r = R^{-1} \quad \forall r = 1, ..., R$. These draws can be generated by different strategies. Classically, a random number generator is

used. Since draws generated by a computer can never be truly random, these draws are often labelled pseudo-random draws. These nodes are often clustered in certain areas. Antithetic sampling algorithms distribute the nodes more evenly but preserve properties of random numbers. In the Monte Carlo experiments, pseudo-random Monte Carlo simulation (PRMC) and antithetic draws from a modified Latin hypercube sampling (MLHS) algorithm are used. It was shown to work well for the estimation of RPL models by Hess, Train, and Polak (2005).
Polynomial cubature methods are multidimensional extensions of Gaussian quadrature. They take a different strategy and determine nodes and weights so that $g(\mathbf{z})$ is approximated by a polynomial of a given order for which the integral is straightforward to solve. In the econometrics literature, one-dimensional Gaussian quadrature methods are known to work well if $g(\mathbf{z})$ is a smooth function and can therefore be well approximated by a (low-order) polynomial (Butler and Moffit 1982). There are different strategies for the generalization of this approach to multiple dimensions. The most straightforward method known as the tensor product rule suffers from the fact that the computational costs rise exponentially with the dimensionality of the problem. It is therefore of little use for more than four or five dimensions. Other methods are not as straightforward to implement and are therefore often considered impractical for the econometric analysis (Bhat 2001, Geweke 1996). The complication lies in the calculation of the nodes $\mathbf{z}_r$ and weights w_r. Given those, they only have to be plugged into equation 5.6. Since the nodes and weights depend only on the dimension of the problem and the desired approximation level, it is also possible to use precalculated values.
Multidimensional cubature methods are derived from one-dimensional Gaussian quadrature formulas. The following discussion is based on the case of fully symmetric weight functions in the sense that the D-variate weighting function w can be multiplicatively decomposed into D univariate functions $\tilde{w}$ as $w(\mathbf{z}) = \tilde{w}(z_1) \cdots \tilde{w}(z_D)$ that are all symmetric so that $\tilde{w}(z_d) = \tilde{w}(-z_d)$ for all $d = 1, \ldots, D$. Most problems in econometrics can be expressed in such a way by a change of variables. See Novak and Ritter (1999) for a discussion of this and more general cases.
In the case of a one-dimensional variable z, the integral in equation 5.5 can be approximated efficiently by Gaussian quadrature methods. Let

$$V_i[g] = \sum_{r=1}^{R(i)} g(z_r^i) w_r^i. \tag{5.7}$$

denote a one-dimensional quadrature rule. The parameter $i \in \mathbb{N}$ drives the precision of this rule. It requires the evaluation of g at a number $R(i)$ of nodes which depends on the intended precision. The nodes $z_1^i, ..., z_{R(i)}^i$ and weights $w_1^i, ..., w_{R(i)}^i$ are given by the quadrature rule. They are constructed such that $V_i[g]$ evaluates the integral exactly with minimum number of function evaluations if g is a polynomial of a certain degree. It is well known that with $R(i) = i$, Gaussian

quadrature rules $V_i[g]$ are able to give exact solutions for all polynomials with an order of at most $2i-1$. The nodes and weights depend on the weight function w and the support Ω. For many standard cases, Gaussian quadrature rules are well known and implemented in many statistical software packages.
In the case of multidimensional $\mathbf{z}$, Gaussian quadrature can be extended by the product rule as discussed by Tauchen and Hussey (1991b). Let $\mathbf{i} = [i_1, ..., i_D]$ denote the vector of precision indices for each dimension. The product rule can be written as

$$T_{D,\mathbf{i}}[g] = (V_{i_1} \otimes \cdots \otimes V_{i_D})[g] \tag{5.8}$$

$$= \sum_{r_1=1}^{R(i_1)} ... \sum_{r_D=1}^{R(i_D)} g(z_{r_1}^{i_1}, ..., z_{r_D}^{i_D}) w_{r_1}^{i_1} \cdots w_{r_D}^{i_D}, \tag{5.9}$$

where the nodes and weights are those implied by the underlying one-dimensional quadrature rules $V_{i_1}, ..., V_{i_D}$. Usually the precision is chosen equally in all dimensions, so the integral is approximated by $T_{D,[i,i,...,i]}[g]$. The curse of dimensionality lies in the fact that the evaluation of this rule requires $R(i)^D$ function evaluations which rises exponentially with D and is prohibitive for high D.
The Smolyak method proposed in this chapter extends Gaussian quadrature rules to multiple dimensions with substantially less function evaluations. This is achieved by combining the univariate rules in a "more clever" way than the product rule. The approach goes back to Smolyak (1963b) and is a general method for multivariate extensions of univariate operators. Integration was already discussed in the original paper and is an active research area in numerical mathematics. Instead of taking the sophisticated one-dimensional rule and naïvely extending it to multiple dimensions by a full product grid, the Smolyak approach is specifically designed for multidimensional problems.
Given an approximation level k, the Smolyak rule linearly combines product rules with different combinations of precision indices $\mathbf{i}$. It can be written as

$$A_{D,k}[g] = \sum_{\mathbf{i} \in S_k^D} (-1)^{D+k-|\mathbf{i}|} \binom{D-1}{D+k-|\mathbf{i}|} T_{D,\mathbf{i}}[g] \tag{5.10}$$

where $S_k^D = \{\mathbf{i} \in \mathbb{N}^D : k+1 \leq |\mathbf{i}| \leq k+D\}$ and $|\mathbf{i}| = i_1 + ... + i_D$. The sum is over all D-dimensional vectors of natural numbers $\mathbf{i}$ which have a norm within certain bounds that are governed by D and k. These vectors translate into the number of nodes for each dimension in a tensor product cubature rule. The bound on the norm has the effect that the tensor product rules with a relatively fine sequence of nodes in one dimension are relatively coarse in the other dimensions.
Equation 5.10 is based on a linear combination of product rules and those in turn are based on one-dimensional Gaussian quadrature rules. Any univariate quadrature rule can serve as a basis for this multivariate extension. A careful choice can further reduce the computational burden. This is true if for some

$\mathbf{i} \in S_k^D$, the product rule $T_{D,\mathbf{i}}$ evaluates g at the same nodes as for other vectors in this set. Instead of repeatedly evaluating the function at the same nodes, this can be done once and only the corresponding weights have to be aggregated. This is possible if the nodes of the one-dimensional basis rules with a low precision index i are also used in those with a high precision index j so that $\{z_1^i, ..., z_{R(i)}^i\} \subset \{z_1^j, ..., z_{R(j)}^j\}$ if $i < j$. For a discussion of this issue also see (Novak and Ritter 1996).

Different rules for generating sets of nodes with this property for Gaussian quadrature are discussed by Petras (2003). In the Monte Carlo experiments for the RPL model shown below, a Smolyak cubature rule based on delayed Kronrod-Patterson sequences as suggested by Petras is used. It is defined for $w(\mathbf{z}) = 1$ and $\Omega = [0,1]^D$. This is adequate for the evaluation of expectations over uniformly distributed random variables and is the extension of a Gauss-Legendre quadrature rule. A problem that frequently occurs is the expectation over standard normal random variables. For this case, Genz and Keister (1996a) apply the Smolyak extension to a Gauss-Hermite quadrature rule, also based on nested sets of nodes, also see Novak and Ritter (1999).

A simple example may help to clarify the approach. Let $D = k = 2$. This implies $S_k^D = \{\mathbf{i} \in \mathbb{N}^2 : 3 \leq |\mathbf{i}| \leq 4\} = \{[1,2],[2,1],[1,3],[3,1],[2,2]\}$. The strategy of Petras (2003) is based on nested sets of nodes. This is achieved with delayed Kronrod-Patterson sequences with $R(1) = 1$, $R(2) = 3$, and $R(3) = 7$. The sets of nodes are $[0.5]$ for $i = 1$, $[0.11, 0.5, 0.89]$ for $i = 2$, and $[0.02, 0.11, 0.28, 0.5, 0.72, 0.89, 0.98]$ for $i = 3$. The set of nodes for lower i are subsets of those with higher i.

For all $\mathbf{i} \in S_k^D$, Figure 5.1 shows the nodes used by the corresponding product rule. Obviously, the nodes for $\mathbf{i} = [1,2]$ and $\mathbf{i} = [2,1]$ are a subset of the nodes for $\mathbf{i} = [1,3]$ and $\mathbf{i} = [3,1]$, respectively. The grid for $\mathbf{i} = [2,2]$ adds four distinct nodes. The complete set of all nodes used by the Smolyak rule only consists of the 17 nodes depicted in the lower right panel of the graph. The full product rule with the corresponding degree of exactness would require the evaluation of the function at the full grid of $7^2 = 49$ nodes. In higher dimensions, this difference becomes more dramatic. With $D = 10$ and $k = 2$, The Smolyak rule needs $1,201$ and the product rule $7^{10} = 282,475,249$ function evaluations.

Monte Carlo integration is very general in the sense that under mild regularity conditions, the approximated integral is $\sqrt{R}$-consistent by a law of large numbers. This convergence rate is independent of the number of dimensions D. However, the error given a finite number R does increase with D. Gauss quadrature and Smolyak cubature rely on the approximation of g by a polynomial. For smooth functions this allows a faster convergence than simulation (Gerstner and Griebel 2003b). For a given number of evaluations, the performance depends on how well g can be approximated by a polynomial of the corresponding order.

Figure 5.1: Construction of the Smolyak grid

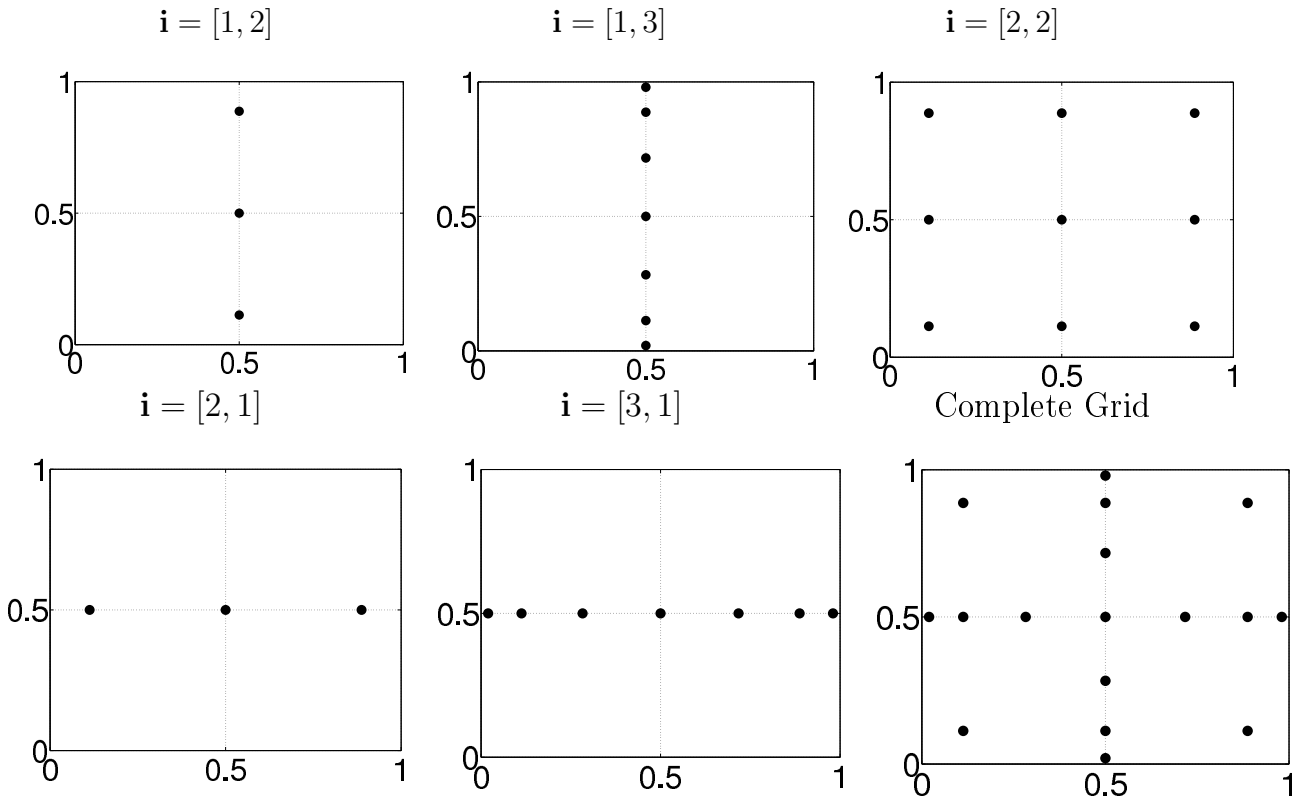


Figure 5.2 shows the performance of different methods in terms of absolute errors for a simple example for which a closed-form expression exists: $\int_{[0,3]^D} \lambda(\mathbf{z})d\mathbf{z}$, where λ is the joint p.d.f. of D i.i.d. logistic random variables. The figure shows the results for $D = 2$ and $D = 5$. Since λ is very smooth, the cubature methods exhibit a much faster convergence rate than simulation. The difference between the tensor product and the Smolyak rule is apparent from a comparison between $D = 2$ and $D = 5$: The higher the dimension, the more efficient the Smolyak is relative to the tensor rule.

5.4 Monte Carlo Experiments

A number of Monte Carlo experiments help to evaluate the relative performance of the numerical integration algorithms. Different random parameters logit (RPL) models as discussed in section 5.2 are implemented. The models are specified for N individuals with T choices between J alternatives each, where each alternative is characterized by K properties. In most model specifications, the K individual taste parameters are normally distributed across the population with mean μ, variance σ^2, and zero covariance. Below also results for uniformly distributed

taste levels are reported in order to test the sensitivity of the results with respect to the specification of the distribution.
As a starting point, a reference model is specified with $N = 1000$, $T = 5$, $J = 5$, $K = 10$, $\mu = 1$, and $\sigma = 0.5$. Then each of these numbers is varied separately to demonstrate their impact on the approximation errors of the different methods. For each of these settings, estimates were obtained for 100 artificial data sets. The properties of the alternatives $\mathbf{x}_{itj}$ were drawn from a standard uniform distribution. The model parameters μ and σ were estimated for each data set.
In order to approximate the integral in equation 5.4 with normally distributed $\boldsymbol{\beta}_i$ by Gauss-Hermite quadrature, it has to be expressed in terms of standard normal random variables. This can be easily done by a change of variables:

$$\int_{\mathbb{R}^K} P_i^*(\boldsymbol{\beta}_i) f_\beta(\boldsymbol{\beta}_i; \boldsymbol{\mu}, \boldsymbol{\sigma})\, d\boldsymbol{\beta}_i = \int_{\mathbb{R}^K} P_i^*\left(\boldsymbol{\mu} + L(\boldsymbol{\sigma})\mathbf{e}_i\right)\right)\, \phi^K(\mathbf{e}_i)\, d\mathbf{e}_i, \tag{5.11}$$

where L is the Cholesky factorization of the covariance matrix of β_i and ϕ^K denotes the joint p.d.f. of K i.i.d. standard normal random variables.
Two simulation-based and two Smolyak-based estimators are implemented and compared in terms of performance. In applications of the RPL models, the maximum simulated likelihood method is predominant for estimation. The pseudo-random Monte Carlo simulation (PRMC) method uses random numbers and the modified Latin hypercube sampling simulation (MLHS) method uses the antithetic quasi random numbers suggested by Hess, Train, and Polak (2005) for the RPL model. In addition, two versions of Smolyak-based cubature for the approximation of the likelihood contributions discussed in section 5.3 are implemented. In the first part, we report the results from the Genz and Keister (1996a) Smolyak cubature (GKSC) rules for the integration over Gaussian distributions, since equation 5.11 has the form required for this method. In addition, the results obtained by the Petras (2003) Smolyak cubature (PSC) rules are reported to test for the sensitivity of the results with respect to the choice of the cubature rule. Both were discussed in section 5.3.
The maximization algorithm for the likelihood function does not affect the relative performance of the estimators based on different numerical integration rules. The standard Newton method with numerically approximated gradients and a BHHH approximation of the Hessian works fine in this application and was used for all estimates.

Reference Model

Figure 5.3 shows the results for the reference model. The performance measure on the ordinate is a relative root mean squared error. Given a certain number of nodes at which the functions are evaluated, the parameters were estimated for 100 simulated data sets using all three methods. For both parameters μ and σ, the mean squared errors were calculated over the 100 replications. They were

then normalized by the respective variance of the best performing method with the maximal number of function evaluations. This makes the MSEs of both parameters comparable between each other and across different model specifications. The number of 100 replications with different data sets is sufficient for conclusive comparisons. In Figure 5.3, the remaining randomness is visualized with error bars indicating the 95% confidence intervals for the respective result. They were generated by resampling from the estimated parameters with replacement and recalculating the performance measure for each of the samples.
Table 5.1 shows the same data, but the results are relative to GKSC with the same number of function evaluations. The confidence intervals take correlations of the results into account. The results are striking. For a large number of function evaluations $R = 1201$, all methods perform equally well. But for moderate and small numbers of evaluations, there are enormous differences. While the Smolyak cubature method needs only 21 evaluation to achieve a negligible approximation error, the error of MLHS and PRMC is higher by a factor of 1.9 and 3.4, respectively. Put differently: The error of GKSC with 21 function evaluations is lower than the error of MLHS with 201 and the error of PRMC with 1201 evaluations.

Varying the Number of Integral Dimensions K

In the following, the results are discussed for similar models in which the parameters of the data generating process are varied. One of the most important parameters is the number K of characteristics defining each of the alternatives, since it equals the dimension of the integral. Figure 5.4 shows the results for the dimensions $K = 4$ and 20. The Smolyak-based GKSC method performs well with a very small number of replications, whereas the simulation-based methods require significantly more computations. A closer look shows that this effect increases somewhat in higher dimensions: While 201 evaluations suffice for MLHS to catch the GKSC performance in 4 dimensions, it does not in ten and the relative difference of PRMC is even higher. The results for other dimensions are qualitatively the same. They are presented together with tables equivalent to table 5.1 in the appendix.

Varying the Variance σ^2

Another important parameter of the data-generating process is σ. The higher its value, the more the integrand varies with the unobserved individual tastes over which the integration is performed. Figure 5.5 shows the results of models with $\sigma = 0.25$ and 1. As expected, all methods perform worse with a higher σ. The relative performances remain similar. With very high σ, all methods fail to give reasonable results.

Other Variations: N, T, J, and μ

The number N of independent observations corresponds to the number of integrals to be solved for each likelihood evaluation. With a higher N, random fluctuations of the approximation error "average out", but a systematic bias remains. This explains the results shown in Figure 5.6 which might be counter-intuitive at first sight: the higher N, the worse do the simulation-based approaches perform relative to Smolyak cubature. With rising N, the variance of the simulation-based estimators decreases, also relative to the sampling variance. But the bias is largely unaffected by N and since the results are expressed in terms of the sampling variance, this drives the MSE up.
Variations of the number of alternatives J or the number of individual choice situations T give similar results. The more data, the higher is the advantage of Smolyak-based cubature over simulation. The parameter μ does not have any impact on the performance of the approximation methods. Loosely speaking, it merely shifts the function to be integrated. These and other results are shown in the appendix.

Taste Distributions and the Cubature Rules

The original formulation of the model with normally distributed taste levels suggests the Genz/Keister cubature method, since it is appropriate for expectations over normally distributed random variables. But the problem can also be restated by another change of variables to become an expectation over standard uniform random variables. The integral in equation 5.11 can be rewritten as

$$\int_{\mathbb{R}^K} P_i^* \left(\boldsymbol{\mu} + L(\boldsymbol{\sigma})\mathbf{e}_i\right)\right) \, \phi^K(\mathbf{e}_i) \, d\mathbf{e}_i = \int_{[0,1]^K} P_i^* \left(\boldsymbol{\mu} + L(\boldsymbol{\sigma})\Phi(\mathbf{u}_i)\right)\right) \, d\mathbf{u}_i, \tag{5.12}$$

where Φ denotes the element-wise standard normal c.d.f. This is the form of integral that the Smolyak cubature method of Petras (2003) requires. The same idea can be applied vice versa: if the model specifies $\boldsymbol{\beta}_i$ to be uniformly distributed, the Petras method can be applied directly and the Genz/Keister method requires a change of variables.
Figure 5.7 compares the performance of both methods with the simulation estimators for these different model specifications. The Genz/Keister rule performs slightly better than Petras in the normal case and slightly worse in the uniform case. But the differences are insignificant and both methods clearly outperform the simulation-based estimators. This evidence suggests that the choice of the Smolyak cubature method is of minor importance for our application.

5.5 Conclusions

Multidimensional integrals are prevalent in econometric estimation problems. Only for special cases, closed-form solutions exist. With a flexible model specification, the researcher frequently has to resort to numerical integration techniques. For one-dimensional integrals, Gaussian quadrature is known to be a powerful tool. Its most straightforward extension to multiple dimensions involves full tensor products of the one-dimensional rules. This implies computational costs that rise exponentially with the number of dimensions, making it infeasible for more than four or five dimensions.

The development of simulation techniques made numerical integration available in general settings. This inspired further development of models for which estimation was previously infeasible. One important example is the mixed or random parameters logit (RPL) model which became one of the most widely used discrete choice models in applied work in recent years. While simulation techniques provide a powerful and flexible approach, they often still require a lot of evaluations of the integrand until the approximation error becomes negligible. Thereby they often impose substantial computational costs.

An intuitive explanation of the advantage of quadrature over simulation in low dimensions is that it efficiently uses the smoothness of the integrand to recover its shape over the whole domain. This chapter proposes a strategy to extend this approach and its advantages to multiple dimensions with dramatically less computational costs than the full tensor rule. It is based on the general method to extend univariate operators to multivariate settings by Smolyak (1963b) which is also appropriate for problems like function approximation.

Extensive Monte Carlo evidence is presented for the RPL model. The results show that the computational costs to achieve a negligible approximation error are much lower with the Smolyak-based approaches than with simulation estimators. Since they only depend on the dimension of the integral and the desired approximation level, the nodes at which to evaluate the integrand and weights can be calculated once or obtained from external sources. Given these, the Smolyak-based methods are straightforward to implement since they only involve calculating weighted averages of integrand values.

Recent research in numerical mathematics suggests possible refinements of the Smolyak approach. First, instead of predefining an approximation level in terms of the number of nodes, a critical value of the approximation error which is easily measured can be specified and the required number of function evaluations can be determined automatically. Second, the approximation does not have to be refined in each dimension symmetrically. It is also possible to invest more effort in the most relevant dimensions. These dimensions can also be determined automatically in an adaptive fashion (Gerstner and Griebel 2003b). Third, quadrature-based methods can handle functions that are not well-behaved for example due to singularities by piecewise integration. These areas can also be determined in

an automated fashion. The exploration of the usefulness of these extensions is left for further research.

Well-behaved integrands are typical in econometric analysis. For these, Smolyak-based cubature provides an efficient and easily applicable alternative to simulation. The efficiency gains can be invested in a reduction of computer time, an improvement of the estimates, and/or a more flexible model specification.

Figure 5.2: Approximation of a Logistic c.d.f. in 5 dimensions

(a) $D = 2$ Dimensions (b) $D = 5$ Dimensions

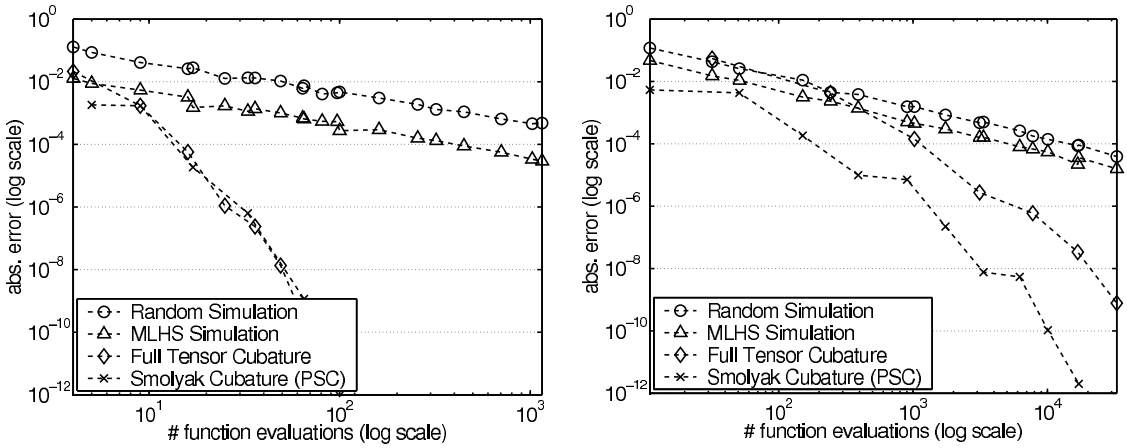


Figure 5.3: Monte Carlo Results: Reference Model

$N = 1000$, $T = 5$, $J = 5$, $K = 10$, $\mu = 1$, $\sigma = 0.5$

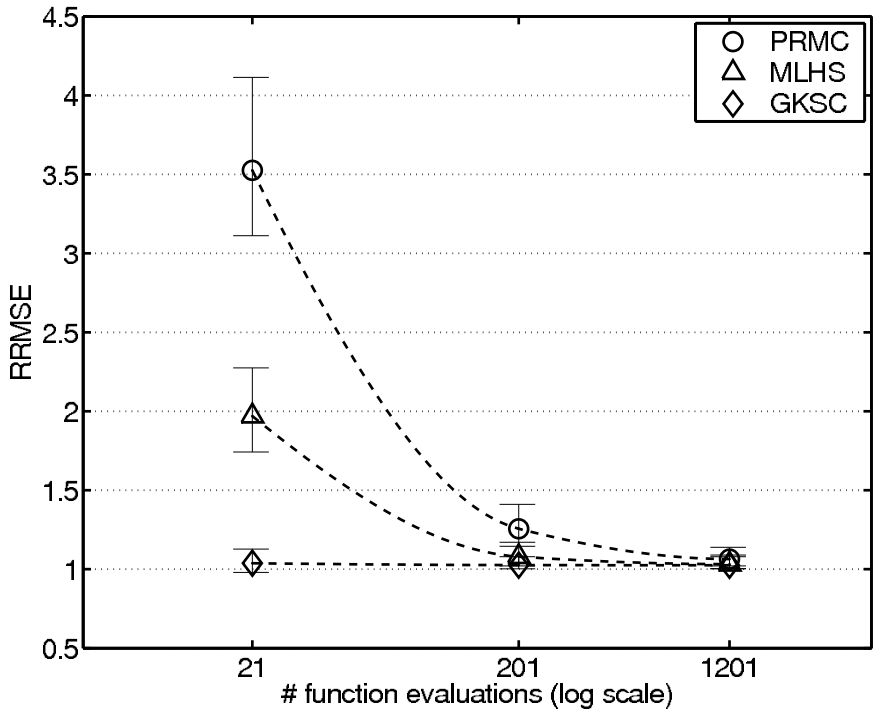


Table 5.1: Monte Carlo Results: Reference model

	$\text{RRMSE}_{\text{GKSC}}$		$\frac{\text{RRMSE}_{\text{PRMC}}}{\text{RRMSE}_{\text{RRMSE}}}$		$\frac{\text{RRMSE}_{\text{MLHS}}}{\text{RRMSE}_{\text{RRMSE}}}$	
R	est.	95% CI	est.	95% CI	est.	95% CI
21	1.04	0.98–1.13	3.40	3.04–3.87	1.90	1.73–2.11
201	1.03	1.00–1.08	1.22	1.13–1.33	1.05	1.01–1.10
1201	1.03	1.00–1.08	1.04	1.01–1.06	1.01	0.99–1.02

Figure 5.4: Monte Carlo Results: Different Dimensions K

All results for: $N = 1000, T = 5, J = 5, \mu = 1, \sigma = 0.5$

(a) $K = 4$ Dimensions (b) $K = 20$ Dimensions

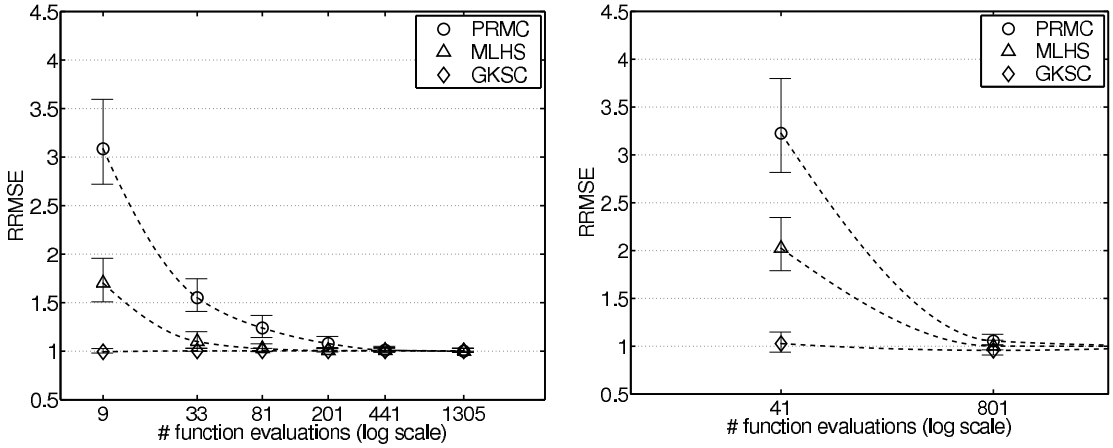


Figure 5.5: Monte Carlo Results, Differences by σ

All results for: $N = 1000, T = 5, J = 5, K = 10, \mu = 1$

(a) $\sigma = 0.25$ (b) $\sigma = 1$

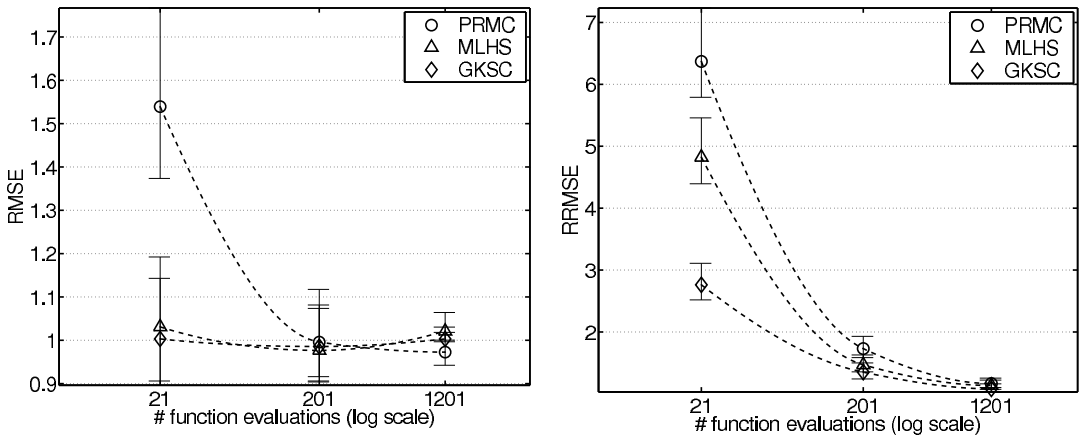


Figure 5.6: Monte Carlo Results, Differences by N

All results for: $T = 5, J = 5, K = 10, \mu = 1, \sigma = 0.5$

(a) $N = 500$ (b) $N = 2000$

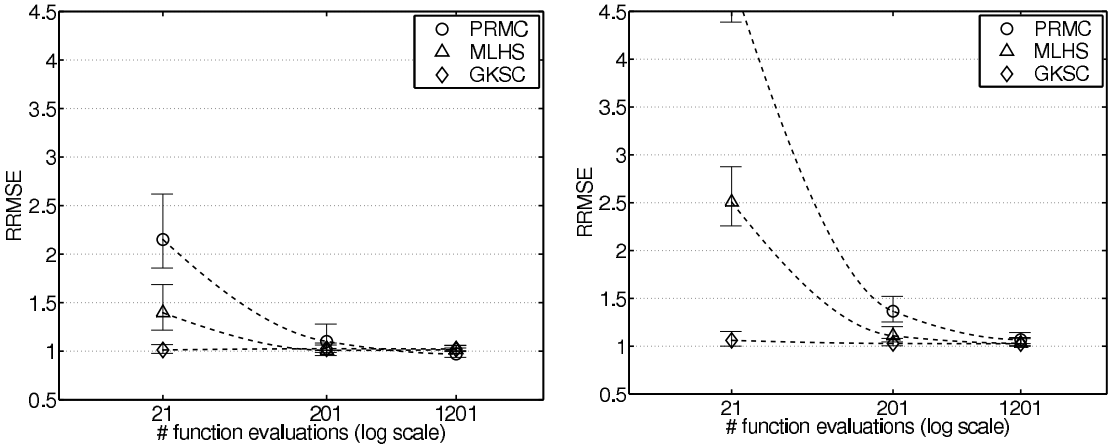
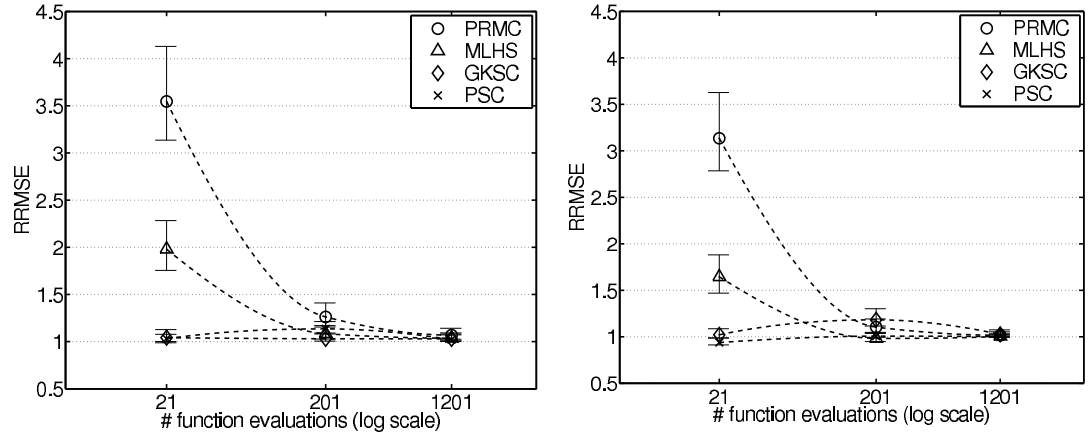


Figure 5.7: Monte Carlo Results: Petras (2003) vs. Genz/Keister (1996)

All results for: $N = 1000, T = 5, J = 5, K = 10, \mu = 1, \sigma = 0.5$

(a) Normal $\boldsymbol{\beta}_i$ (b) Uniform $\boldsymbol{\beta}_i$



5.6 Appendix

This appendix presents further results mentioned but not shown in the main text.

Table 5.2: Monte Carlo Results: Differences by K: 2–10

	$\text{RRMSE}_{\text{GKSC}}$		$\frac{\text{RRMSE}_{\text{PRMC}}}{\text{RRMSE}_{\text{RRMSE}}}$		$\frac{\text{RRMSE}_{\text{MLHS}}}{\text{RRMSE}_{\text{RRMSE}}}$	
R	est.	95% CI	est.	95% CI	est.	95% CI
Dimension $K = 2$						
5	1.01	0.99–1.05	2.67	2.38–3.05	1.85	1.63–2.12
9	1.01	1.00–1.05	2.31	2.03–2.67	1.22	1.12–1.35
17	1.01	1.00–1.05	1.66	1.49–1.89	1.07	1.01–1.14
45	1.01	1.00–1.05	1.25	1.14–1.40	1.03	1.01–1.07
401	1.01	1.00–1.05	1.02	0.99–1.06	1.00	0.99–1.01
961	1.01	1.00–1.05	1.00	0.98–1.02	1.00	0.99–1.00
Dimension $K = 4$						
9	0.99	0.98–1.03	3.10	2.72–3.60	1.71	1.52–1.95
33	1.00	1.00–1.03	1.55	1.40–1.72	1.10	1.03–1.18
81	1.00	1.00–1.03	1.23	1.13–1.35	1.02	0.99–1.06
201	1.00	1.00–1.03	1.08	1.02–1.13	1.00	0.99–1.02
441	1.00	1.00–1.03	1.01	0.98–1.03	1.00	0.99–1.02
1305	1.00	1.00–1.03	1.00	0.98–1.01	1.00	0.99–1.00
Dimension $K = 6$						
13	0.98	0.96–1.02	3.03	2.67–3.53	1.52	1.39–1.70
73	1.00	1.00–1.03	1.38	1.26–1.52	1.03	0.98–1.09
257	1.00	1.00–1.03	1.07	1.01–1.14	1.01	0.99–1.02
749	1.00	1.00–1.03	1.02	0.99–1.05	1.00	0.99–1.01
2021	1.00	1.00–1.03	1.01	0.99–1.02	1.00	0.99–1.00
Dimension $K = 8$						
17	1.00	0.97–1.05	3.64	3.24–4.14	1.87	1.69–2.10
129	1.00	1.00–1.03	1.35	1.25–1.47	1.00	0.95–1.06
609	1.00	1.00–1.03	1.04	1.01–1.08	1.00	0.98–1.01
2193	1.00	1.00–1.03	1.01	0.99–1.02	1.00	0.99–1.01
Dimension $K = 10$						
21	1.04	0.98–1.12	3.40	3.04–3.89	1.90	1.72–2.12
201	1.03	1.00–1.08	1.22	1.14–1.33	1.05	1.01–1.10
1201	1.03	1.00–1.08	1.04	1.01–1.06	1.01	0.99–1.02

Table 5.3: Monte Carlo Results: Differences by K: 12–20

	$\text{RRMSE}_{\text{GKSC}}$		$\frac{\text{RRMSE}_{\text{PRMC}}}{\text{RRMSE}_{\text{RRMSE}}}$		$\frac{\text{RRMSE}_{\text{MLHS}}}{\text{RRMSE}_{\text{RRMSE}}}$	
R	est.	95% CI	est.	95% CI	est.	95% CI
Dimension $K = 12$						
25	0.96	0.90–1.03	3.36	3.07–3.74	1.94	1.78–2.14
289	1.02	1.00–1.05	1.10	1.00–1.19	0.97	0.94–1.02
2097	1.00	1.00–1.03	1.00	0.98–1.02	1.00	0.99–1.02
Dimension $K = 14$						
29	0.96	0.91–1.03	3.28	2.95–3.73	1.86	1.67–2.10
393	1.01	0.99–1.04	1.03	0.95–1.10	0.98	0.92–1.03
3361	1.00	1.00–1.02	1.00	0.98–1.02	1.01	1.00–1.03
Dimension $K = 16$						
33	0.96	0.89–1.07	3.29	2.93–3.74	1.97	1.79–2.20
513	1.00	1.00–1.03	1.05	0.96–1.14	1.00	0.95–1.06
Dimension $K = 18$						
37	1.04	0.93–1.19	3.66	3.29–4.16	2.14	1.95–2.41
649	1.00	1.00–1.03	1.03	0.97–1.10	0.98	0.92–1.04
Dimension $K = 20$						
41	1.03	0.94–1.15	3.14	2.82–3.52	1.97	1.82–2.15
801	0.96	0.91–1.01	1.10	1.05–1.15	1.05	0.99–1.10
10001	1.00	1.00–1.02	1.00	0.98–1.03	1.00	0.97–1.03

Table 5.4: Monte Carlo Results: Differences by σ

	$\text{RRMSE}_{\text{GKSC}}$		$\frac{\text{RRMSE}_{\text{PRMC}}}{\text{RRMSE}_{\text{RRMSE}}}$		$\frac{\text{RRMSE}_{\text{MLHS}}}{\text{RRMSE}_{\text{RRMSE}}}$	
R	est.	95% CI	est.	95% CI	est.	95% CI
$\sigma = 0.25$						
21	1.00	0.89–1.14	1.53	1.38–1.72	1.03	0.94–1.14
201	0.99	0.92–1.08	1.01	0.94–1.09	0.99	0.96–1.02
1201	1.00	1.00–1.03	0.97	0.93–1.01	1.02	0.99–1.04
$\sigma = 0.5$						
21	1.04	0.98–1.12	3.40	3.04–3.89	1.90	1.72–2.12
201	1.03	1.00–1.08	1.22	1.14–1.33	1.05	1.01–1.10
1201	1.03	1.00–1.08	1.04	1.01–1.06	1.01	0.99–1.02
$\sigma = 1$						
21	2.76	2.52–3.11	2.31	2.23–2.39	1.75	1.69–1.80
201	1.35	1.24–1.50	1.28	1.22–1.35	1.09	1.03–1.14
1201	1.08	1.03–1.16	1.08	1.04–1.12	1.05	1.01–1.09

Table 5.5: Monte Carlo Results: Differences by N

	$\text{RRMSE}_{\text{GKSC}}$		$\frac{\text{RRMSE}_{\text{PRMC}}}{\text{RRMSE}_{\text{RRMSE}}}$		$\frac{\text{RRMSE}_{\text{MLHS}}}{\text{RRMSE}_{\text{RRMSE}}}$	
R	est.	95% CI	est.	95% CI	est.	95% CI
$N = 200$						
21	0.95	0.84–1.09	1.44	1.28–1.64	1.10	1.01–1.21
201	0.99	0.95–1.05	0.98	0.92–1.05	0.98	0.93–1.04
1201	1.00	1.00–1.03	0.99	0.95–1.02	0.97	0.96–0.99
$N = 500$						
21	1.01	0.98–1.07	2.12	1.83–2.57	1.38	1.20–1.64
201	1.02	1.01–1.06	1.07	0.96–1.23	0.98	0.93–1.04
1201	1.02	1.01–1.06	0.95	0.91–1.00	0.99	0.98–1.01
$N = 1000$						
21	1.04	0.98–1.12	3.40	3.04–3.89	1.90	1.72–2.12
201	1.03	1.00–1.08	1.22	1.14–1.33	1.05	1.01–1.10
1201	1.03	1.00–1.08	1.04	1.01–1.06	1.01	0.99–1.02
$N = 2000$						
21	1.06	1.00–1.15	4.58	4.16–5.13	2.36	2.17–2.63
201	1.03	1.00–1.08	1.33	1.24–1.42	1.08	1.03–1.13
1201	1.03	1.01–1.08	1.04	1.02–1.07	1.00	0.99–1.02

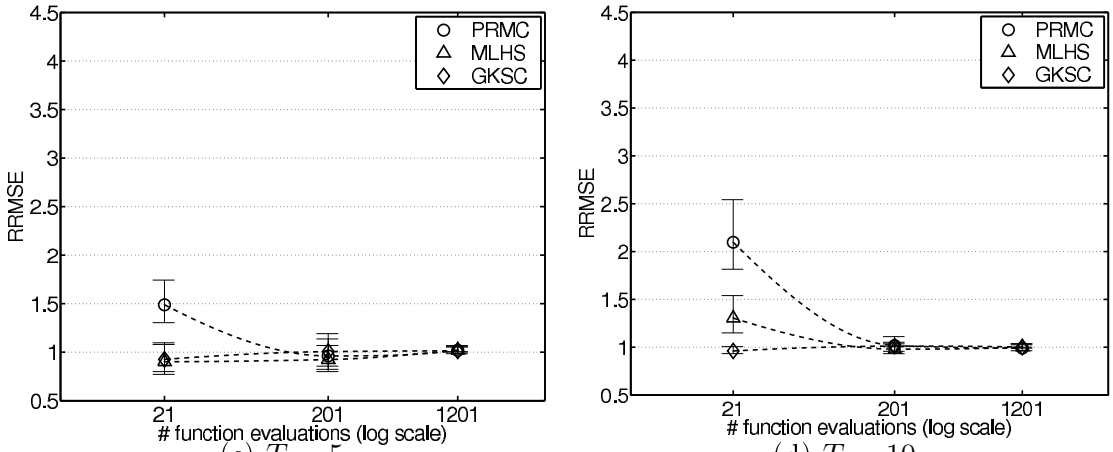
Table 5.6: Monte Carlo Results: Differences by T

	$\text{RRMSE}_{\text{GKSC}}$		$\frac{\text{RRMSE}_{\text{PRMC}}}{\text{RRMSE}_{\text{RRMSE}}}$		$\frac{\text{RRMSE}_{\text{MLHS}}}{\text{RRMSE}_{\text{RRMSE}}}$	
R	est.	95% CI	est.	95% CI	est.	95% CI
$T = 2$						
21	0.93	0.80–1.10	1.61	1.40–1.85	0.97	0.87–1.10
201	1.01	0.86–1.19	0.96	0.89–1.04	0.92	0.87–0.98
1201	1.01	1.00–1.05	1.01	0.97–1.05	1.01	0.99–1.03
$T = 3$						
21	0.96	0.93–1.01	2.18	1.88–2.64	1.36	1.20–1.59
201	1.01	1.00–1.04	1.01	0.94–1.09	0.97	0.92–1.03
1201	1.00	1.00–1.03	0.99	0.96–1.03	1.00	0.99–1.02
$T = 5$						
21	1.04	0.98–1.12	3.40	3.04–3.89	1.90	1.72–2.12
201	1.03	1.00–1.08	1.22	1.14–1.33	1.05	1.01–1.10
1201	1.03	1.00–1.08	1.04	1.01–1.06	1.01	0.99–1.02
$T = 10$						
21	1.37	1.27–1.52	3.32	2.94–3.72	2.04	1.81–2.27
201	1.31	1.21–1.46	1.12	1.04–1.21	1.04	0.99–1.09
1201	1.31	1.22–1.46	0.99	0.98–1.01	0.99	0.98–1.01

Figure 5.8: Monte Carlo Results, Differences by T

All results for: $N = 1000, J = 5, K = 10, \mu = 1, \sigma = 0.5$

(a) $T = 2$ (b) $T = 3$



(c) $T = 5$ (d) $T = 10$

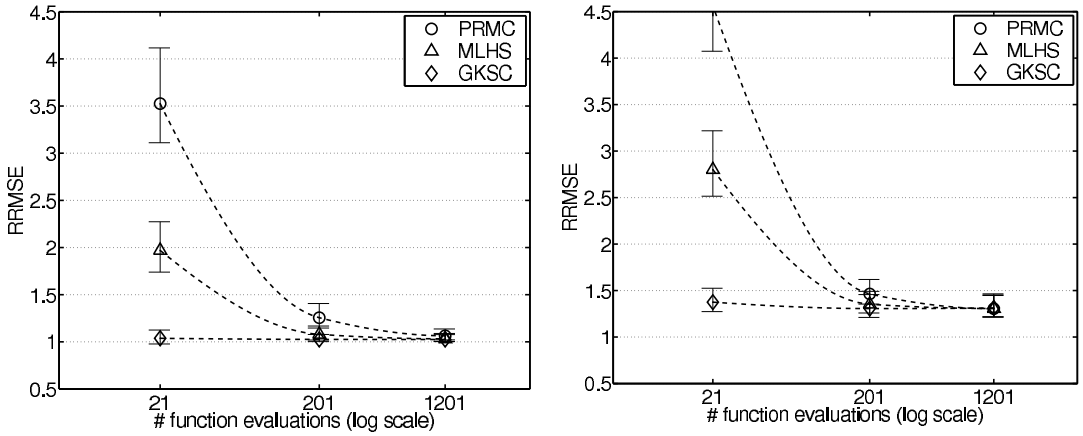
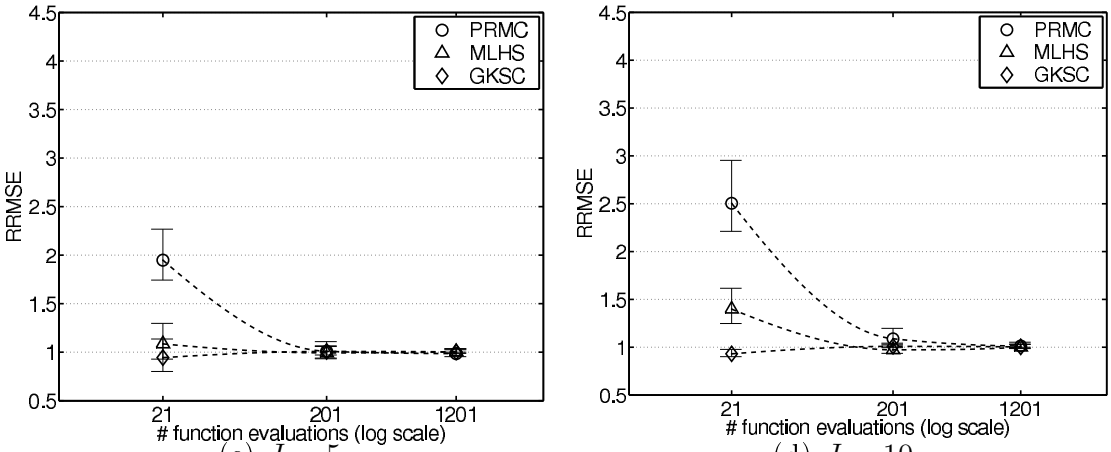


Figure 5.9: Monte Carlo Results, Differences by J

All results for: $N = 1000, T = 5, K = 10, \mu = 1, \sigma = 0.5$

(a) $J = 2$ (b) $J = 3$



(c) $J = 5$ (d) $J = 10$

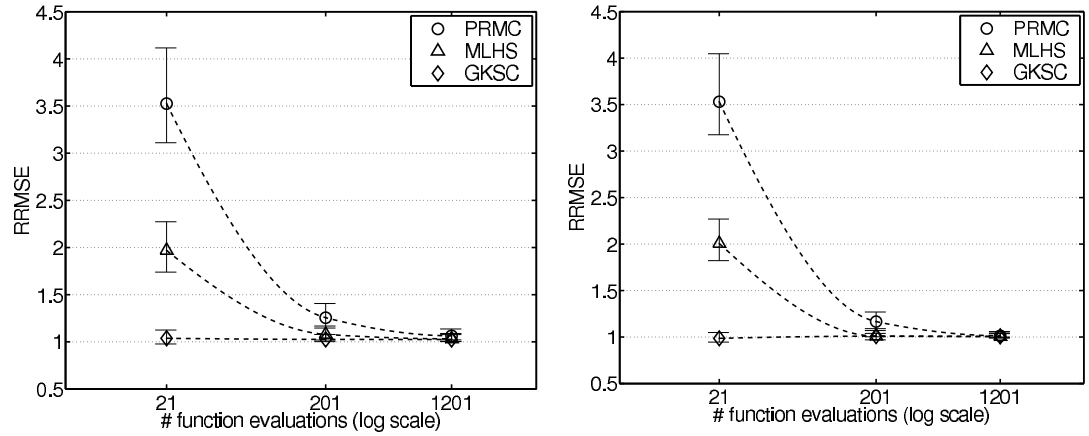


Table 5.7: Monte Carlo Results: Differences by J

	$\text{RRMSE}_{\text{GKSC}}$		$\frac{\text{RRMSE}_{\text{PRMC}}}{\text{RRMSE}_{\text{RRMSE}}}$		$\frac{\text{RRMSE}_{\text{MLHS}}}{\text{RRMSE}_{\text{RRMSE}}}$	
R	est.	95% CI	est.	95% CI	est.	95% CI
$J = 2$						
21	0.94	0.80–1.14	2.06	1.82–2.37	1.15	1.03–1.29
201	1.01	0.97–1.06	1.00	0.92–1.09	0.99	0.92–1.05
1201	1.00	1.00–1.03	0.98	0.95–1.02	1.00	0.98–1.02
$J = 3$						
21	0.93	0.90–0.98	2.69	2.36–3.17	1.50	1.34–1.72
201	1.01	1.00–1.04	1.08	0.99–1.18	0.96	0.91–1.02
1201	1.00	1.00–1.03	1.01	0.98–1.04	1.00	0.98–1.02
$J = 5$						
21	1.04	0.98–1.12	3.40	3.04–3.89	1.90	1.72–2.12
201	1.03	1.00–1.08	1.22	1.14–1.33	1.05	1.01–1.10
1201	1.03	1.00–1.08	1.04	1.01–1.06	1.01	0.99–1.02
$J = 10$						
21	0.99	0.95–1.05	3.58	3.22–4.05	2.04	1.87–2.25
201	1.01	1.00–1.04	1.16	1.08–1.24	1.00	0.96–1.04
1201	1.01	1.00–1.04	1.01	0.99–1.03	1.00	0.99–1.01

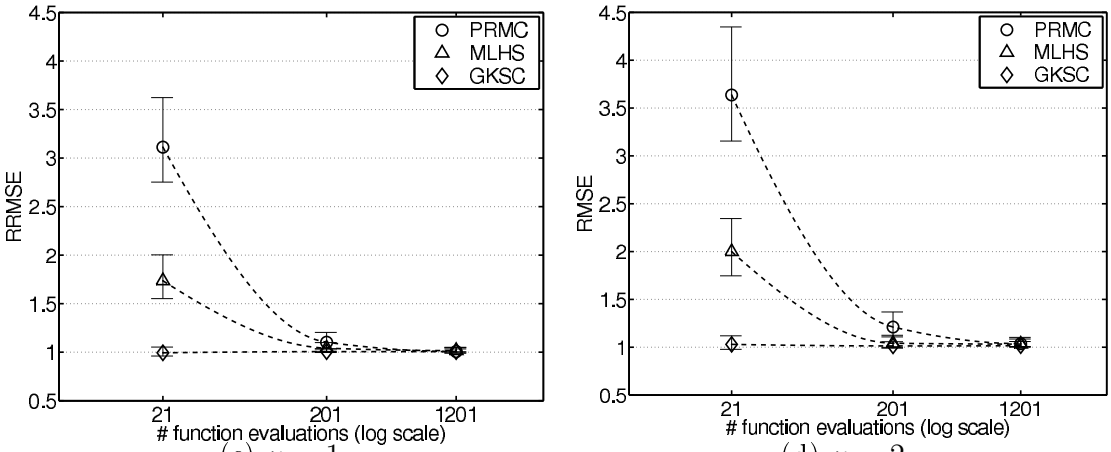
Table 5.8: Monte Carlo Results: Differences by μ

	$\text{RRMSE}_{\text{GKSC}}$		$\frac{\text{RRMSE}_{\text{PRMC}}}{\text{RRMSE}_{\text{RRMSE}}}$		$\frac{\text{RRMSE}_{\text{MLHS}}}{\text{RRMSE}_{\text{RRMSE}}}$	
R	est.	95% CI	est.	95% CI	est.	95% CI
$\mu = 0$						
21	0.99	0.96–1.05	3.13	2.79–3.57	1.75	1.58–1.96
201	1.01	1.00–1.04	1.10	1.03–1.17	1.03	1.00–1.07
1201	1.01	1.00–1.03	1.00	0.99–1.02	1.01	1.00–1.02
$\mu = 0.5$						
21	1.03	0.98–1.12	3.53	3.16–4.00	1.94	1.77–2.15
201	1.01	1.00–1.06	1.19	1.10–1.30	1.03	0.99–1.07
1201	1.02	1.00–1.06	1.02	1.00–1.05	1.02	1.00–1.03
$\mu = 1$						
21	1.04	0.98–1.12	3.40	3.04–3.89	1.90	1.72–2.12
201	1.03	1.00–1.08	1.22	1.14–1.33	1.05	1.01–1.10
1201	1.03	1.00–1.08	1.04	1.01–1.06	1.01	0.99–1.02
$\mu = 2$						
21	0.98	0.94–1.04	3.29	2.96–3.71	1.79	1.61–2.00
201	1.01	1.00–1.04	1.25	1.13–1.39	1.02	0.97–1.06
1201	1.00	1.00–1.04	1.05	1.01–1.10	1.00	0.98–1.02

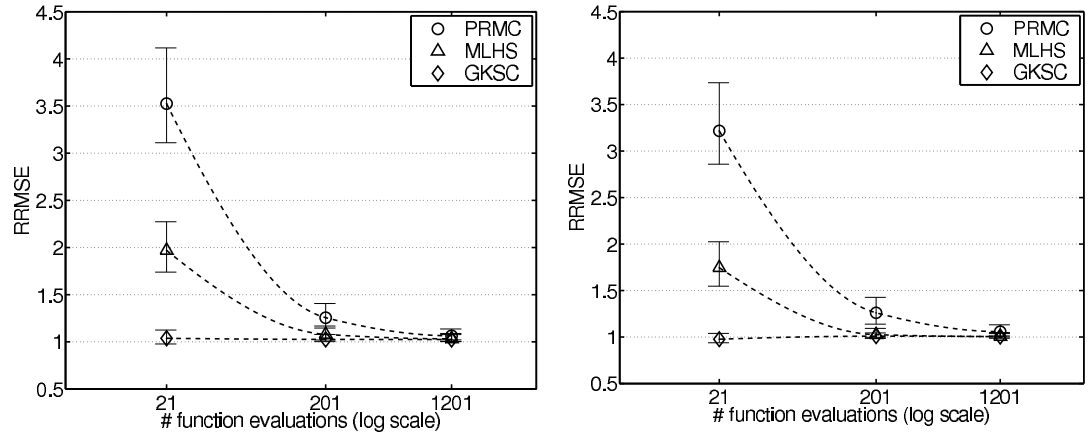
Figure 5.10: Monte Carlo Results, Differences by μ

All results for: $N = 1000, T = 5, J = 5, K = 10, \sigma = 0.5$

(a) $\mu = 0$ (b) $\mu = 0.5$



(c) $\mu = 1$ (d) $\mu = 2$



Bibliography

(1998): "Blinded by the dark," *The Economist*, p. 13.

Aruoba, B. S., J. Fernádez-Villaverde, and J. F. Rubio-Ramírez (2003): "Comparing Solution Methods for Dynamic Equilibrium Economies," Discussion paper, University of Pennsylvania.

Bayoumi, T. (1992): "The Effect of the ERM on Participating Economies," *International Monetary Fund Staff Papers*, 39(2), 331–356.

Bayoumi, T., and B. Eichengreen (1994): "One Money or Many? Analysing the Prospects for Monetary Unification in Various Parts of the World," *Princeton Studies in International Finance*.

Belsley, D. A., E. Kuh, and R. E. Welsch (1980): *Regression diagnostics : identifying influential data and sources of collinearity*. John Wiley, New York.

Berger, J., and R. Wolpert (1988): *The Likelihood Principle*. Hayward, California, 2^{nd} edn.

Bhat, C. (2001): "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," *Transportation Research B*, 35, 677–693.

Blanchard, O. J., and C. M. Kahn (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48, 1305–1312.

Blanchard, O. J., and D. Quah (1989): "The Dynamic Effects Of Aggregate Demand And Supply Disturbances," *American Economic Review*, 79(4), 655–673.

Börsch-Supan, A., and V. Hajivassiliou (1993): "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models," *Journal of Econometrics*, 58, 347–368.

Brooks, S. P., and A. Gelman (1998): "General methods for monitoring convergence of iterative simulations.," *Journal of Computational & Graphical Statistics*, 7(4), 434–455.

Brown-Humes, C., and C. MacCarthy (2000): "Fear of identity loss puts result on knife-edge," *Financial Times, 28 Sep.*

Butler, J. S., and R. Moffit (1982): "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," *Econometrica*, 50(3), 761–764.

Chadha, B., and E. Prasad (1997): "Real exchange rate fluctuations and the business cycle," *International Monetary Fund Staff Papers*, 44, 328–355.

Chib, S., and E. Greenberg (1995): "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49(4), 327–335.

Clarida, R., and J. Gali (1994): "Sources of Real Exchange-Rate Fluctuations: How Important are Nominal Shocks?," *Carnegie-Rochester Series on Public Policy*, 41, 1–56.

Dibooğlu, S., and A. M. Kutan (2000): "Sources of Real Exchange Rate Fluctuations in Transition Economies: The Case of Poland and Hungary," Discussion Paper B14- 00, Zentrum für Europäische Integrationsforschung.

Doornik, J. A. (2002): *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press and Oxford, London, 3rd edn.

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

Eckstein, Z., and K. Wolpin (1999): "Why Youths Drop out of High School: The Impact of Preferences, Opportunities, and Abilities," *Econometrica*, 67, 1295–1339.

Eichengreen, B. (1992): "Should the Maastricht Treaty be Saved?," *Princeton Studies in International Finance*.

Enders, W., and B.-S. Lee (1997): "Accounting for real and nominal exchange rate movements in the post-Bretton Woods period," *Journal of International Money and Finance*, 16(2), 233–254.

FAZ GmbH Informationsdienste (ed.) (1998): *Mittel- und Osteuropa Perspektiven, Jahrbuch 1998/99*, vol. 2. Frankfurter Allgemeine Zeitung, Frankfurt am Main.

Fernádez-Villaverde, J., and J. F. Rubio-Ramírez (2004a): "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood," Discussion paper, University of Pennsylvania.

——— (2004b): "Estimating Nonlinear Dynamic Equilibrium Economies: A Likelihood Approach," Discussion paper, University of Pennsylvania.

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2004a): "Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach," *Journal of Econometrics*, 123, 153–187.

Frankel, J. A., and A. Rose (1999): "An Estimate of the Effect of Currency Unions on Trade and Growth," Discussion Paper 2631, Centre for Economic Policy Research.

Frankel, J. A., and A. K. Rose (1998): "The Endogeneity of the Optimum Currency Area Criteria," *The Economic Journal*, 108(449), 1009–1025.

Friedman, M. (1953): *Essays in Positive Economics*. University of Chicago Press, Chicago and London.

Friedman, M., and L. Savage (1948): "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, 56, 279–304.

——— (1952): "The Expected Utility Hypothesis and the Measurability of Utility," *Journal of Political Economy*, 60, 463–474.

Gelfand, A., and D. Dey (1994): "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of the Royal Statistical Society, Series B*, 56, 501–514.

Gelman, A., and D. B. Rubin (1992): "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7(4), 457–472.

Genz, A. (1986): "Fully Symmetric Interpolatory Rules for Multiple Integrals," *SIAM Journal on Numerical Analysis*, 23, 1273–1283.

Genz, A., and B. D. Keister (1996a): "Fully Symmetric Interpolatory Rules for Multiple Integrals Over Infinite Regions with Gaussian Weight," *Journal of Computational and Applied Mathematics*, 71, 299–309.

Genz, A., and C. Keister (1996b): "Fully symmetric interpolatory rules for multiple intergrals over infinite regions with Gaussian weights," *Journal of Computational and Applied Mathematics*, 71, 299–309.

Gerstner, T., and M. Griebel (2003a): "Dimension–Adaptive Tensor–Product Quadrature," *Computing*, 71(1), 65–87.

——— (2003b): "Dimension-Adaptive Tensor-Product Quadrature," *Computing*, 71, 65–87.

GEWEKE, J. (1996): "Monte Carlo Simulation and Numerical Integration," in *Handbook of Computational Economics Vol. 1*, ed. by H. M. Amman, D. A. Kendrick, and J. Rust, pp. 731–800. Elsevier Science, Amsterdam.

——— (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18, 1–126.

——— (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, New York.

GORDON, N. J., D. J. SALMOND, AND A. F. M. SMITH (1993): "A novel approach to nonlinear and non-Gaussian Bayesian state estimation," *IEE Proceedings on Radar and Signal Processing*, 140, 107–113.

GOURIÉROUX, C., AND A. MONFORT (1996): *Simulation-Based Econometric Methods*. Oxford University Press.

GRAUWE, P. D. (1997): *The economics of monetary integration*. Oxford Univiversity Press, Oxford, 3. edn.

HAJIVASSILIOU, V. A., AND P. A. RUUD (1994): "Classical Estimation Methods for LDV Models Using Simulation," in *Handbook of Econometrics Vol. IV*, ed. by R. F. Engle, and D. L. McFadden, pp. 2383–2441. Elsevier, New-York.

HANSEN, L. P., AND J. J. HECKMAN (1996): "The Empirical Foundations of Calibration," *Journal of Economic Perspectives*, 10(1), 87–104.

HARVEY, A. (1985): "Trends and cycles in macroeconomic time series," *Journal of Business and Economic Statistics*, 3, 216–227.

HASSELBLATT, C., I. PROOS, AND V. ZIRNASK (1996): "Estland," in *Mittel- und Osteuropa auf dem Weg in die Europäische Union*, ed. by W. Weidenfeld. Verlag Bertelsmann Stiftung, 2. edn.

HEINEMANN, F. (1995): "Central Europe and European Monetary integration, A Strategy for Catching up," *Intereconomics*.

HEISS, F., AND V. WINSCHEL (2005): "Smolyak Cubature for Multiple Integration in Estimation Problems," Discussion paper.

HESS, S., K. TRAIN, AND J. POLAK (2005): "On the Use of a Modified Latin Hypercube Sampling (MLHS) Method in the Estimation of a Mixed Logit Model for Vehicle Choice," *Transportation Research B*, forthcoming.

JUDD, K. (1992): "Projection methods in economics," *Journal of Economic Theory*, 58, 410–452.

JUDD, K. L. (1998): *Numerical Methods in Economics*. MIT Press, Cambridge, Mass.

——— (2002): "Perturbation Methods with Nonlinear Changes of Variables," Discussion paper, Hoover Institution, Stanford.

JULIER, S. J., AND J. K. UHLMANN (1997): "A new extension of the Kalman filter to nonlinear systems," in *The proceedings of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls, Orlando, Florida*.

——— (2002): "The Scaled Unscented Transformation," in *Proceedings of the IEEE American Control Conference*, pp. 4555–4559.

——— (2004): "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, pp. 401–422.

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.

KENEN, P. B. (1969): "The Theory of Optimum Currency Areas: An Eclectic View," in *Monetary Problems of the International Economy*, ed. by R. A. Mundell, and A. K. Swoboda, pp. 41–60. University of Chicago Press, Chicago and London.

KIM, J., AND S. KIM (2003): "Spurious welfare reversals in international business cycle models," *journal of International Economics*, 60(2), 471–500.

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.

KLIMKE, A. (2004): "Multilinear sparse grid interpolation in matlab," Discussion paper, University of Stuttgart.

KOLLMANN, R. (2002): "Monetary Policy Rules in the Open Economy: Effects on Welfare and Business Cycles," *Journal of Monetary Economics*, 49, 989–1015.

——— (2003): "Monetary Policy Rules in an Interdependent World," Discussion Paper 4012, CEPR Discussion Paper.

KOTECHA, J., AND P. DJURIĆ (2003a): "Gaussian Particle Filter," *IEEE Transactions on Signal Processing*, 51, 2592–2601.

——— (2003b): "Gaussian Sum Particle Filter," *IEEE Transactions on Signal Processing*, 51, 2602–2612.

Krüger, D., and F. Kübler (2004): "Computing equilibrium in OLG models with stochastic production," *Journal of Economic Dynamics and Control*, 28, 1411–1436.

Krugman, P. (1993): "Lessons of Massachusetts for EMU," in *The Transition to Economic and Monetary Union in Europe*, ed. by F. Giavazzi, and F. Torres. Cambridge University Press, New York.

Landon-Lane, J. S. (1998): "Bayesian Comparison of Dynamic Macroeconomic Models," Ph.D. thesis, University of Minnesota.

Marimon, R., and A. Scott (eds.) (1999): *Computational Methods for the Study of Dynamic Economies*. Oxford university Press, Oxford.

McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Choice Models Without Numerical Integration," *Econometrica*, 57, 995–1026.

McFadden, D., and K. Train (2000): "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 447–470, Unveröffentlichtes Manuskript, University of California, Berkeley.

McGrattan, E. (1998): "Application of weighted residual methods to dynamic economic models," Discussion Paper Staff Report 232, Federal Reserve Bank of Minneapolis.

McKinnon, R. I. (1963): "Optimum Currency Areas," *American Economic Review*, 53, 717–724.

Miranda, M. J., and P. L. Fackler (2002): *Applied Computational Economics and Finance*. MIT Press, Cambridge, MA.

Morley, J., and J. Piger (2005): "The Importance of Nonlinearity in Reproducing Business Cycle Features," Discussion paper, The Federal Reserve Bank of St.Louis.

Mundell, R. A. (1961): "A Theory of Optimum Currency Areas," *American Economic Review*, 51, 657–664.

Muth, J. F. (1961): "Rational Expectations and the Theory of Price Movements," *Econometrica*, 29, 315–335.

Novak, E., and K. Ritter (1996): "High Dimensional Integration of Smooth Functions over Cubes," *Numerische Mathematik*, 75, 79–97.

Novak, E., and K. Ritter (1999): "Simple cubature formulas with high polynomial exactness," *Constructive Approximation*, 15, 499–522.

Obstfeld, M. (1985): "Floating Exchange Rates: Experience and Prospects," *Brookings Papers on Economic Activity*, 0, 369–450.

Obstfeld, M., and K. Rogoff (1995): "Exchange rate dynamics redux," *Journal of Political Economy*, 103(3), 624–660.

Patterson, T. (1968): "The Optimum Addition of Points to Quadrature Formulae," *Mathematics of Computation*, 22, 847–856.

Pearl, J. (2000): *Causality*. Cambridge, University Press.

Petras, K. (1999): "Fast Calculation of Coefficients in the Smolyak Algorithm," Discussion paper, Technische Universität Braunschweig.

Petras, K. (2003): "Smolyak Cubature of Given Polynomial Degree with Few Nodes for Increasing Dimension," *Numerische Mathematik*, 93, 729–753.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1988): *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, Massachusetts.

Ricci, L. A. (1997): "A Model of an Optimum Currency Area," Discussion Paper WP/97/76, International Monetary Fund.

Rose, A. (2000): "One money, one market: estimating the effect of common currencies on trade," *Economic Policy*, 15(30).

Rotemberg, J. J., and M. Woodford (1998): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy: Expanded Version," Discussion Paper 233, National Bureau of Economic Research.

Ruge-Murcia, F. J. (2003): "Methods to Estimate Dynamic Stochastic General Equilibrium Models," Discussion paper, Université de Montréal.

Schmitt-Grohé, S., and M. Uribe (2004): "Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function," *Journal of Economic Dynamics and Control*, (28), 645–658.

Smolyak, S. (1963a): "Quadrature and interpolation formulas for tensor products of certain classes of functions," *Soviet Math.Dokl*, 4, 240–243.

Smolyak, S. A. (1963b): "Quadrature and Interpolation Formulas for Tensor Products of Certain Classes of Functions," *Soviet Mathematics Doklady*, 4, 240–243.

STORN, R., AND K. PRICE (1995): "Differential Evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces," Discussion Paper TR-95-012, International Computer Science Institute,, Berkeley.

TAUCHEN, G., AND R. HUSSEY (1991a): "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models," *Econometrica*, 59(2), 371–396.

——— (1991b): "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models," *Econometrica*, 59(2), 371–396.

TAYLOR, J. B. (1993): *Macroeconomic Policy in a World Economy: From Econometric Design to Practical Operation*. Norton, New York, London.

TER BRAAK, C. J. (2004): "Genetic algorithms and Markov Chain Monte Carlo: Differential Evolution Markov Chain makes Bayesian computing easy," Discussion paper, Biometris Research Centre, Wageningen University.

TRAIN, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.

VAN DER MERWE, R., N. DE FREITAS, A. DOUCET, AND E. WAN (2001): "The Unscented Particel Filter," in *Advances in Neural Information Processing Systems 13*.

VAUBEL, R. (1976): "Real exchange rate changes in the European Community: the empirical evidence and its implication for the European currency unification," *Weltwirtschaftliches Archiv*, 112, 429–470.

WALSH, C. E. (2001): *Monetary Theory and Policy*. MIT Press, Cambridge, Massachusetts.

WASILKOWSKI, G. W., AND H. WOŹNIAKOWSKI (1999): "Weighted tensor product algorithms for linear multivariate problems," *Journal of Complexity*, 15, 402–447.

Ehrenwörtliche Erklärung

Hiermit erkläre ich ehrenwörtlich, daß ich diese Dissertationsschrift selbständig angefertigt habe und mich anderer als der in ihr angegebenen Hilfsmittel nicht bedient habe. Entlehnungen aus anderen Schriften sind ausdrücklich als solche gekennzeichnet und mit Quellenangaben versehen.

Mannheim, 10. Oktober 2005

Viktor Winschel

Lebenslauf

Viktor Winschel

Universität Mannheim
L7,3-5
68131 Mannheim

Zur Person

Geboren am 6.6.1970 in Duschanbe, UdSSR

Aktuelle Tätigkeit

Seit Okt. 2005 Mitarbeiter am Lehrstuhl Prof. Felix Kübler, Ph.D.

Werdegang

2000-2005	Promotion in VWL bei Prof. Dr. Roland Vaubel
1994-2000	Diplom an der Universität Mannheim
1998-1999	Auslandsstudium an der Université Catholique de Louvain, Belgien
1992-1994	Gründung and Management der media mutant Filmproduktion, Stuttgart
1991-1992	Zivildienst
1990	Abitur