

**Moment-Based Estimation of Macroscopic
Dynamic Models in Macroeconomics and Finance**

ALS INAUGURAL-DISSERTATION
ZUR ERLANGUNG DES AKADEMISCHEN GRADES EINES DOKTORS
DER WIRTSCHAFT- UND SOZIALWISSENSCHAFTLICHEN FAKULTÄT
DER CRISTIAN-ALBRECHTS-UNIVERSITÄT ZU KIEL

vorgelegt von

M. A. Tae-Seok Jang,
aus Seoul (Korea) geb. 29/01/1979

Kiel, November 2012

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

Dekan: Prof. Horst Raff
Erstberichterstattender: Prof. Dr. Thomas Lux
Zweitberichterstattender: Prof. Dr. Roman Liesenfeld

Tag der Abgabe der Arbeit: 28 August, 2012
Tag der mündlichen Prüfung: 31 October, 2012

Acknowledgments

This research has been supported by the German Academic Exchange Service (DAAD) and the University of Kiel. The author greatly acknowledges his academic mentor and research supervisors, Reiner Franke, Thomas Lux and Roman Liesenfeld for their guidance and support; he is also indebted to Stephen Sacht for his collaborative work on the DSGE models. He appreciates all comments and suggestions that he received at conferences in Berlin, Kiel, Mannheim, Zwickau (Germany), Dallas (US), London (UK), Prague (Czech Republic), and Seoul (Korea). He is grateful to his family, friends, and colleagues in the world.

Citation Guides

Main body of my PhD thesis will contain five chapters. Note here that the thesis includes collaborative works with Reiner Franke and Stephen Sacht. Citations to these works can be done as follows:

Chapter 2 Backward-Looking Behavior in the New-Keynesian Model

FRANKE, R., AND JANG, T.-S. AND SACTH, S. (2011): *Moment matching versus Bayesian estimation: backward-looking behaviour in the New-Keynesian three-equations model*. Economic Working paper 2011-10, University of Kiel.

Chapter 3 Structural Estimation of the New-Keynesian Model

JANG, T.-S. (2012): Structural Estimation of the New-Keynesian Model: a Formal Test of Backward- and Forward-looking Behavior, in: N. Balke, F. Canova, F. Milani, and M. Wynne (ed.), *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Development*, *Advances in Econometrics: Vol. 28*, Emerald Group Publishing Limited, 421–467.

Chapter 4 Identification of Animal Spirits in a Bounded Rationality Model

JANG, T.-S. AND SACTH, S. (2012): *Identification of animal spirits in a bounded rationality model: an application to the Euro area*. Economic Working paper 2012-12, University of Kiel.

Chapter 5 The Role of Behavioral Heterogeneity in Asset Pricing Models

JANG, T.-S. (2012): *The role of behavioral heterogeneity in asset pricing models*. mimeo, University of Kiel.

Chapter 6 Social Interaction Effects in Financial Data

JANG, T.-S. (2012): *Identification of social interaction effects in financial data: simulation based inference for a noise trader model*. mimeo, University of Kiel.

Abstract

The thesis is divided into two parts. Part I (macroeconomic models) is devoted to empirically examine the role of backward-looking behavior in a standard New-Keynesian model and its behavioral variant. This part includes three chapters. In chapter 2, the structural parameters of the New-Keynesian model (NKM) are estimated from a historical data set of the US economy. The moment-matching method is used to discuss the importance of backward-looking behavior in the New-Keynesian Phillips Curve, and its empirical results are contrasted to the Bayesian estimation. Then the model with purely forward-looking expectations and its hybrid variant are compared using a formal test. Chapter 3 discusses the persistent dynamics of inflation and output in the NKM, and analyzes statistical properties of the estimation methods and the choice of moment conditions. The empirical performance of model selection methods is examined using information criteria along the lines of the maximum likelihood estimation. Chapter 4 demonstrates that market euphoria and depression in the economy can be explained by switching between heterogeneous investors along the lines of the discrete choice theory. This chapter investigates a bounded rationality model with historical Euro Area data and discuss the effects of the investors' over(or under)reaction on market structure. Part II (financial market models) contains two chapters concerning the relevance of behavioral heterogeneity in financial markets. Chapter 5 discusses the effect of heterogeneous trading rules on the return volatility in the asset pricing model. The adaptive belief system is estimated using the simulated method of moments estimator. Especially, two types of the noise term in the model dynamics are investigated by means of simulations (i.e. additive and multiplicative). In a structural stochastic volatility model, the two trading mechanisms (i.e. wealth and herding) are compared according to a formal test. Finally, chapter 6 examines the social interaction effects of a market microsimulation model on various historical FX data. In particular, simulation based inference is used to examine the validity of the group behavior when the analytical expression for moment conditions is fairly complicated.

Keywords: behavioral heterogeneity; group behavior; moment-based estimation; microsimulation; New-Keynesian; return volatility; social interaction effects.

Contents

List of Figures	V
List of Tables	VII

1 Introduction	1
1.1 Motivation of the study	1
1.2 Structure of the thesis	5

PART I. Macroeconomic Models

2 Backward-Looking Behavior in the New-Keynesian Model	7
2.1 Introduction	7
2.2 The moment matching estimation approach	9
2.3 The three-equations model	14
2.4 The Great Inflation period	16
2.4.1 Basic results	16
2.4.2 Price indexation versus shock persistence	19
2.4.3 Is full price indexation significantly superior?	23
2.4.4 Admitting stronger backward-looking behavior	28
2.4.4 Evaluation of the estimated parameters	31
2.5 The Great Moderation period	34
2.6 Conclusion	40
Appendix to chapter 2	
A Prior densities of the Bayesian reference estimation	42
B The standard percentile and Hall's percentile confidence interval	
.....	43
C Estimation of the moment covariance matrix $\hat{\Sigma}_m$	43

D Specification of the matrices V^X and V^Y	46
3 Structural Estimation of the New-Keynesian Model	49
3.1 Introduction	49
3.2 Expectation formation in a DSGE model	51
3.2.1 The New-Keynesian three-equations model	52
3.2.2 Sources of persistence: backward- and forward-looking behavior	54
3.3 Estimation methodologies and model selection	55
3.3.1 Method of moment and model comparison	55
3.3.2 Maximum likelihood and model selection	58
3.4 Empirical application	60
3.4.1 Data	60
3.4.2 Basic results on method of moments estimation and model comparison	60
3.4.3 Basic results on maximum likelihood estimation	65
3.4.4 Validity of extra moment conditions	67
3.5 Attaining efficiency from moment conditions	73
3.5.1 Monte Carlo study	73
3.5.2 Model selection and discussion	76
3.6 Conclusion	78
Appendix to chapter 3	
A Choice of moments	79
B Reduced form of matrix and solution of the NKM	80
C VAR lag order selection	81
D Matrix notation	82
E Technical note on the model comparison method	83
F Simulated QLR distribution for model comparison	85

G The Monte Carlos result of the misspecified case	87
4 Identification of Animal Spirits in a Bounded Rationality Model	89
4.1 Introduction	89
4.2 The model: rational expectations vs. bounded rationality	91
4.3 The estimation methodology	96
4.4 Empirical application to the Euro area	100
4.4.1 Data	100
4.4.2 Basic results	100
4.4.3 Comparison with other studies	107
4.5 Robustness checks	108
4.5.1 Monte Carlo study	108
4.5.1 Sensitivity of the behavioral parameters	110
4.5 Conclusion	114
Appendix to Chapter 4	
A Solution of the NKM	116
B Delta method and confidence interval for auto- and cross-covariances	118
C Large-scale simulation study for the BR model	119

PART II. Financial Market Models

5 The Role of Behavioral Heterogeneity in Asset Pricing Models	121
5.1 Introduction	121
5.2 The asset pricing model with behavioral heterogeneity	123
5.2.1 The adaptive belief system	124
5.2.2 A structural stochastic volatility model	130

5.3 The estimation methodology	132
5.4 Empirical application	134
5.4.1 Data and choice of moments	134
5.4.2 Estimation results of ABS	135
5.4.3 Estimation results of SSV and model comparison	138
5.5 Conclusion	142
Appendix to Chapter 5	
A Stochastic interpretation of switching rule in ABS	143
B Selection matrix	144
C Technical note on the model comparison method	145
D Double moving block bootstrap method	147
E Simulated QLR distribution for model comparison	148
6 Social Interaction Effects in Financial Data	149
6.1 Introduction	149
6.2 The model framework	151
6.2.1 The behavior of noise traders	151
6.2.2 The artificial market and benchmark simulation	155
6.3 Simulation based inference	157
6.3.1 The simulated method of moments estimation	158
6.3.2 The moment conditions and the block bootstrap method	159
6.3.3 The algorithm and the hypothesis test	161
6.3.4 Collinearity and parameter identification	164
6.4 Empirical applications	168
6.4.1 Data	168
6.4.1 Parameter calibration I: system size and herding tendency	169
6.4.2 Parameter calibration II: trading volume	170
6.4.3 Decomposition of return volatility	171

6.5 Conclusion	174
Appendix to Chapter 6	
A Choice of Moments	176
B Summary Statistics of the FX Data	177
7 General Conclusion	181
References	185
Curriculum Vitae	209

List of Figures

2.1	Estimated <i>versus</i> empirical covariance profiles (GI).	19
2.2	Minimized values of J given α and ρ_π (GI).	20
2.3	MM estimation of the model under exogenous variations of α (GI).	22
2.4	Covariance profiles of model F from Table 2.3 (bold lines).	31
2.5	Frequency distributions of the re-estimations of the bootstrapped model A (GI).	33
2.6	Frequency distributions of the re-estimations of the bootstrapped model A (GM).	37
3.1	Covariance profiles for inflation persistence in GI (dashed: empirical, Δ : hybrid, *: forward)	69
3.2	Covariance profiles for inflation persistence in GM (dashed: empirical, Δ : hybrid, *: forward)	70
3.3	Test distribution for inflation persistence: GI (left) and GM (right)	85
3.4	Test distribution for output persistence: GI (left) and GM (right)	85
3.5	Test distribution for inflation persistence: GI (left) and GM (right)	86
3.6	Test distribution for output persistence: GI (left) and GM (right)	86
4.1	Dynamics in the output gap and the inflation gap.	104
4.2	Model covariance (Cov) profiles in the Euro Area.	105
4.3	3-D contour plot of the parameter space with β and δ	112
4.4	3-D contour plot of the parameter space with β and ρ	113
4.5	3-D contour plot of the parameter space with δ and ρ	113
5.1	Deterministic structure of ABS	127
5.2	Return series of ABS with additive (left) and multiplicative (right) noise.	129
5.3	Distribution of the bootstrapped objective function values	135
5.4	The effect of noise specification on objective function: additive (left) and multiplicative (right).	137
5.5	Returns (upper) and fraction of chartists (lower) from herding	140
5.6	Returns (upper) and fraction of chartists (lower) from wealth	141
5.7	Deterministic model dynamics with a finite number of agents	143
5.8	QLR distribution of SSV	148
6.1	The opinion index and its frequency histogram	153
6.2	The behavior of switching process in a noise trader model	154
6.3	Noise trader model with social interactions	157
6.4	Prices (upper) and returns (lower) of historical DM/US	160
6.5	Hill plot for daily returns	161

6.6	Three-dimensional contour plot of objective function	163
6.7	Block Bootstrap	164
6.8	3-D contour plots	166
6.9	The behavior of objective function w.r.t the parameter ν	168
6.10	Model-generated data using the estimated parameters	173
6.11	Prices (upper) and returns (lower) of CHF/US data	177
6.12	Prices (upper) and returns (lower) of FRF/US data	178
6.13	Prices (upper) and returns (lower) of GBP/US data	179
6.14	Prices (upper) and returns (lower) of JPY/US data	180

List of Tables

2.1	Parameter estimates for GI.	17
2.2	Comparison of alternative estimations.	27
2.3	GI estimations (the economic constraints on α and χ are dropped).	29
2.4	Pairwise correlations of parameter re-estimates of model A (GI).	34
2.5	Parameter estimates for GM.	35
2.6	Pairwise correlations of parameter re-estimates of model A (GM).	38
2.7	Prior densities of the BR estimations in Tables 2.1 and 2.5.	42
2.8	Specification of the index sets.	45
2.9	The last three columns of matrix \hat{G}	46
3.1	Sources of persistence in the NKPC and the IS equation	54
3.2	Parameter estimates for inflation persistence with 15 moments	61
3.3	Empirical and model-generated moments for inflation persistence: 15 moment conditions	62
3.4	Parameter estimates for output persistence with 15 moments	63
3.5	Empirical and model-generated moments for output persistence: 15 moment conditions	64
3.6	ML estimates for inflation and output persistence	65
3.7	Parameter estimates for inflation persistence with 42 moments	68
3.8	Parameter estimates for output persistence with 42 moments	71
3.9	The Monte Carlo results on the MM and ML estimates, (): root mean square error, S.E : mean of standard error	75
3.10	Model selection using information criteria: inflation persistence	76
3.11	Model selection using information criteria: output persistence	77
3.12	VAR lag order selection criteria	81
3.13	Monte Carlo results on the MM and ML estimates of the misspecified model, (): root mean square error, S.E : mean of standard errors	87
4.1	Estimates of the RE and BR model	101
4.2	Monte Carlo study for the RE model	109
4.3	Monte Carlo study for the BR model	111
4.4	Monte Carlo Study for the BR Model	119
5.1	Parameter values for simulation	126
5.2	Empirical and bootstrapped moment estimates of S&P 500 index	134
5.3	Parameter estimates for ABS	136
5.4	Simulated moments of ABS with additive and multiplicative noise	136
5.5	Parameter estimates of herding and wealth	139
5.6	Simulated moments of herding and wealth mechanisms	140
6.1	Parameter values for simulation	156

6.2	Statistics of empirical moments and block bootstrap samples: DM/US	161
6.3	Grid points in three-dimensional contour plot	165
6.4	Flatness and multimodality I of objective function	167
6.5	Flatness and multimodalities II of objective function	167
6.6	Variation in parameters estimates from system size: DM/US	169
6.7	Parameter estimates I of the noise trader model for FX data	170
6.8	Variation in parameter estimates from trading volume: DM/US	171
6.9	Parameter estimates II of the noise trader model for FX data	172
6.10	Decomposition of stochastic processes for FX data	172
6.11	Moment estimates of FX data	174
6.12	Statistics of empirical moments and block bootstrap samples: CHF/US	177
6.13	Statistics of empirical moments and block bootstrap samples: FRF/US	178
6.14	Statistics of empirical moments and block bootstrap samples:GBP/US	179
6.15	Statistics of empirical moments and block bootstrap samples: JPY/US	180

Introduction

1.1 Motivation of the study

The purpose of this thesis is to empirically examine a standard macroeconomic model and alternative models of behavioral heterogeneity (i.e. agent-based models (ABM)) in finance. At the beginning, we study the structural dynamics of these models and their complexity, some of which do not permit a straightforward application to data. Therefore our empirical analysis aims to demonstrate how to disentangle the structural modeling dynamics and endogeneity from which we can find an empirical connection between the behavior of economic agents and the macroscopic dynamics; i.e. behavioral heterogeneity, social interaction effects, optimal behavior and expectation formation process.¹ Then we develop econometric or statistical inference techniques such that the empirical performance of models can be compared.

Concerning the specific models examined in this thesis, we first discuss a small scale dynamics stochastic general equilibrium (DSGE) model and its behavioral variant in which agents have cognitive limitations. In particular, we attempt to estimate the structural parameters of this kind of macroeconomic model using US and Euro Area data. In the second part, the asset pricing models with behavioral heterogeneity are put forward to explain the effect of heterogeneous trading rules on financial markets; we have delved into empirical questions about price movements using the adaptive belief system and a structural stochastic volatility model. Finally, we study the role of group behavior in a market microsimulation model and identify the effects of social interactions on return volatility.

For the most part, this thesis resolves the parameter estimation of the structural models by using moment-based estimation. In this study, we use the terminology of the moment-based estimation (or the minimum distance estimator), which can encompass the generalized method of moments and the classical minimum distance estimator as special cases. Also see Newey and McFadden (1994) for their excellent classification of these econometric estimators. Here we briefly summarize the econometric steps for estimating the DSGE model and ABMs. It is assumed that the

¹The standard macroeconomic model is based on the microeconomic theory of utility/profit maximization where agents discount their future consumption/interest rate under perfect foresight; e.g. consumption smoothing in Euler equations or Calvo price-setting rules. On the other hand, ABMs follow a bottom-up approach using simple decision rules of agents, which can allow for group interactions and behavioral heterogeneity.

p -dimensional set of parameters of the models must satisfy the following moment restrictions:

$$E[m(x_i; \theta_0)] = 0 \quad (1.1)$$

where $m(x_i; \theta)$ is a q -dimensional vector of functions of the observable data x_i , $i = 1, \dots, T$. The model-generated moment conditions will satisfy the moment restriction in Equation (1.1) if the parameter θ in Θ converges to the true one θ_0 . For the parameter identification in the moment-based estimation, our empirical applications consider at least as many moment conditions as parameters: over or exact identification, i.e. $q \geq p$.

In empirical applications, we estimate the sample moment equation:

$$m^* = \frac{1}{T} \sum_{i=1}^T m(x_i) \quad (1.2)$$

where m^* is the chosen empirical moments, which can comprise distributional properties of the empirical data x_i . The objective function of the method of moments estimator is defined as the following quadratic form:

$$S(\theta) = (m^* - m)' W (m^* - m) \quad (1.3)$$

where m is the model-generated moment function. W is a non-negative definite, symmetric weight matrix. Accordingly, we obtain the parameter estimates from the objective function S at its minimum:

$$\hat{\theta} = \arg \min_{\theta \in \Theta} S(\theta) \quad (1.4)$$

Under the standard regularity conditions,² the estimator has the following asymptotic distribution of the parameter estimates:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \sim N(0, \Lambda) \quad (1.5)$$

²The regularity conditions such as stochastic equicontinuity and asymptotic unbiasedness for moment estimates are used to provide consistency, asymptotic normality and efficiency of econometric estimators. A set of the regularity conditions includes the parameter compactness, a twice continuously differentiable objective function and its uniform convergence, etc. See Newey and McFadden (1994) for large sample theory in econometrics.

where Λ is a covariance matrix of the parameter estimation. In general, it is shown in econometric literature that the method of moments estimator is consistent and asymptotically normal; e.g. Hansen (1982), Pakes and Pollard (1989), and Duffie and Singleton (1993). If the model complexity and underlying data generating process hinder accurate assessment of asymptotic inference (e.g. closed-form or analytical solutions are not available), we rely on simulation techniques and bootstrap methods for the estimation.

However, there are pros and cons to apply the moment-based estimation techniques to the complex (or overly parameterized) models. On the one hand, from the model-generated moments, we can transparently evaluate the performance of the models to data and find a good approximation to the data generating process. This can be considered a direct result of the solution of a system instead of relying on an auxiliary model, which indirectly measures an approximation of the model to the data.³ On the other hand, because of the non-linear mapping between the model structure and the reduced form parameters, however, the estimated parameters can be poorly determined with large standard errors; the large sampling variability makes the empirical analysis inherently difficult in the context of statistical inference.

In many cases, complexity hinders a direct estimation approach such as likelihood-based statistical inference, and the objective function exhibits several optima in the parameter space of the complex model; e.g. see Winker et al. (2007). Moreover, the moment conditions are not linear in the parameter vector θ , so that we cannot prove the global identification conditions for the structural parameters; a standard rank condition - a necessary and sufficient condition for the identification of the parameters of structural equations models - cannot be applied. This means that we do not know the existence of a unique optimum in the objective function. The worst case could be that the objective function cannot be differentiated and its convexity is not known.⁴ This has practical implications. For example, the convergence of the optimization procedure should be checked during the estimation. We need to examine a unique optima from different starting values of the model parameters; it is often the case that variations in the parameter estimates should be examined using a set of different optimization routines such as iterative minimization, Nelder-Mead simplex, random search method, etc.

From the current problem in empirical applications, we address an issue of moment-matching approach to evaluating the complex models. But the statistical properties of the estimator by mathematical proofs is beyond the scope of this

³The auxiliary model in indirect inference can refer to a linear regression model, a finite impulse response model, a score vector of efficient method of moments, etc.

⁴The differentiability of the objective function is a mild (or weak) requirement for regularity, since consistency and asymptotic normality of the moment-based estimation can be established with a non-smooth objective function; e.g. see the estimation of discrete choice models in McFadden (1989).

present investigation and is relegated to future research. Therefore the main contribution of this thesis is to take economically well formulated models to data, where we show how to find the solution of the system and to implement some numerical and technical methods such as simulation techniques and comparison procedures between two competing specifications. In other words, we evaluate the empirical performance of the DSGE models and ABMs based on their ability to match the sample moments of data. From these analyses, we attempt to establish an empirical connection between the behavioral rules and the macroscopic dynamics in these models.

In this thesis, we consider that the model is possibly misspecified to capture the reality; i.e. the empirical model is based on an incomplete probability formulation of the real data. In this case, statistical hypothesis testing from the likelihood function is not straightforward; the assumption that error terms follow a Gaussian distribution does not hold. In addition to common statistical problems when the model is misspecified, the closed form solutions for likelihood function often do not exist. Therefore, our purpose of this empirical analysis is to test whether the alternative explanations of the macroscopic dynamics can provide a reasonable approximation to the real data according to the chosen moment conditions. We consider the behavioral heterogeneity in the structural dynamics as a possibly relevant explanation on the data generating process, and discuss to what extent the moment-based estimation can answer the economic questions concerning the behavior of the economic agents.

More generally, the statistical inference procedures adopted in this thesis are as follows: first, we derive analytic moment conditions for a system of equations, and utilize them when matching the empirical properties of data; the stylized facts known in empirical macroeconomic and financial literature (e.g. first moment, the sample auto- and cross-covariances, fat-tailed distribution, etc.) are used for the purpose of parameter estimation and model validation. Second, we judge the model validity based on its ability to match the chosen empirical moments. However, if the moment conditions do not have a simple analytic form, then we replace them by an approximation based on simulations. During the model estimation, we study two competing specifications and compare their empirical performances using a formal test; i.e. hypothesis testing of an equal fit of the two models. Finally, we investigate the properties of the moment-based estimator *via* Monte Carlo experiments. Overall we evaluate the empirical performance of the macroeconomic and financial market models along the lines of moment conditions discussed in this thesis.

1.2 Structure of the thesis

The thesis is divided into two parts. Part I (macroeconomic models) is devoted to empirically examine the role of backward-looking behavior in a standard New-Keynesian model and its behavioral variant. This part includes three chapters. Chapter 2 attempts to estimate the structural parameters of the New-Keynesian model (NKM); e.g. the price indexation parameter. A historical data set of the US economy is used to discuss the importance of backward- or forward-looking behavior in the New-Keynesian Phillips Curve by using both moment-matching and Bayesian estimation. Then the model with purely forward-looking expectations and its hybrid variant are compared using the formal model comparison test of Hnatkovska, Marmer and Tang (2012). Chapter 3 investigates the persistent dynamics of inflation and output in the NKM and provides a formal discussion on the choice of moment conditions for the model estimation. The empirical performance of model selection methods is examined using information criteria along the lines of the maximum likelihood estimation. The similarity between the two chapters is that we use the same specification of a small-scale NKM for the empirical analysis. But a major stress of chapter 3 is on the robustness of the chosen moment conditions and their statistical relationship to the maximum likelihood estimation. Chapter 4 demonstrates that market euphoria and depression in the economy can be explained by switching between heterogeneous groups along the lines of the discrete choice theory. This chapter estimates the behavioral parameters of a bounded rationality model of De Grauwe (2011) with historical Euro Area data by moment-based estimation. The model provides insights for understanding the investors' over(or under)reaction to observed economic shocks and their effects on market structure.

Part II (financial market models) contains two chapters investigating the relevance of behavioral heterogeneity in financial markets. Chapter 5 discusses the effect of heterogeneous trading rules on the return volatility in the asset pricing model. First, the original version of the adaptive belief system (Gaunersdorfer and Hommes (2007)) is estimated using the simulated method of moments estimator. Especially, the model dynamics are investigated by means of simulations with two types of the noise term (i.e. additive and multiplicative). Second, the properties of a structural stochastic volatility model are examined following Franke and Westerhoff (2012). The two trading mechanisms (i.e. wealth and herding) are compared according to a formal test. Finally, chapter 6 discusses the social interaction effects of a market microsimulation model in various historical FX data (Alfarano and Lux (2007)). In particular, simulation based inference is used to examine the validity of the group behavior when the analytical expression for moment conditions is fairly complicated.

A general conclusion summarizes the main findings of the thesis and presents an agenda for future research.

PART I. Macroeconomic Models

Backward-Looking Behavior in the New-Keynesian Model

The chapter considers an elementary New-Keynesian three-equations model and contrasts its Bayesian estimation with the results from the method of moments (MM), which seeks to match the model-generated second moments of inflation, output and the interest rate to their empirical counterparts. Special emphasis is placed on the degree of backward-looking behavior in the Phillips curve. While, in line with much of the literature, it only plays a marginal role in the Bayesian estimations, MM yields values of the price indexation parameter close to or even at its maximal value of one. These results are worth noticing since the matching thus achieved is entirely satisfactory. The matching of some special (and even better) versions of the model is econometrically evaluated by a model comparison test.

2.1 Introduction

The New-Keynesian modelling of dynamic stochastic general equilibrium (DSGE) with its nominal rigidities and incomplete markets is still the ruling paradigm in contemporary macroeconomics. The fundamental three-equations versions represent the so-called New Macroeconomic Consensus and, as a point of departure, are most valuable in shaping the theoretical discussion on monetary policy and other topics. Over the last decade these models have also been extensively subjected to estimation. Here system estimations (as opposed to single-equations estimations) gained in importance. First maximum likelihood and more recently the Bayesian estimation approach crystallized as the most popular methods, a development that probably not the least was fostered by the dissemination of the powerful DYNARE software package. By now Bayesian estimations have even become so dominant that other techniques are at risk of eking out a marginal existence.

The exclusiveness of likelihood methods is nevertheless worth reconsidering. In some form or another, it is well-known that “maximum likelihood does the ‘right’ efficient thing if the model is true. It does not necessarily do the ‘reasonable’ thing for ‘approximate’ models” (Cochrane, 2001, p. 293). This remark, which certainly carries over to the marginal likelihood in the Bayesian estimations, should not be neglected since after all, any model in economics can only be an approximation to the hypothetical construct of a true real-world data generation process. For this reason it is desirable, unless vital, to work with alternative system estimation methods as well.

While likelihood methods concentrate on predictions of a model for the next period, the method of moments (MM) estimation approach, as we understand this term here, is concerned with the dynamic properties of a model in general. Their quantitative representation refers to a number of summary statistics, or ‘moments’, and the estimation seeks to identify numerical parameter values such that the model-generated moments come as close as possible to their empirical counterparts.

The crucial point of MM is obviously the choice of these moments, which by critics is branded as arbitrary.¹ Again, however, the approximate nature of structural modelling must be taken into account, which at best allows a model to match *some* of the ‘stylized facts’ of an actual economy. Limited-information methods like MM are therefore not necessarily inferior to a full-information estimation approach. Moreover, MM requires the researcher to make up his or her mind about the dimensions along which the model should be most realistic, and it is just this explicitness and, in practice, easy interpretation of the moment matching that are strong arguments in favour of MM. This begins informally with diagrams comparing the profiles of the theoretical to the empirical moments and their inspection with the naked eye, but also more formal methods are available to assess a model’s goodness-of-fit. In fact, learning in these ways which of the empirical moments are more, and which are less adequately matched can provide useful information about the particular merits and demerits of a model.

The present chapter takes a New-Keynesian three-equations model from the shelf and contrasts its Bayesian estimations with the results from MM estimations. As far as we know, such a direct comparison has not been undertaken before. Specifically, we start out from the Bayesian estimations of a version that enabled Castelnovo (2010) to demonstrate the superiority of a positive and time-varying inflation target over a steady state rate of inflation fixed at zero. Our interest is, however, more elementary, which is the reason why we circumvent this issue by having the structural equations directly referring to the deviations of inflation and the interest rate from an exogenous trend. We rather concentrate on the sources of inflation persistence in the Phillips curve as they are caused by exogenous or endogenous factors, i.e., by serial correlation in the shock process or by price indexation of firms, where the latter yield a positive coefficient on lagged inflation and a corresponding reduction of the coefficient on expected inflation.

In this respect, Castelnovo in line with several other examples in the literature obtains evidence for strong forward-looking behavior (low indexation) and high correlation in the random shocks. This feature is once again confirmed by the Bayesian estimations of our slightly modified model. By contrast, to anticipate our most important finding, the MM estimations show a strong tendency towards the opposite: high price indexation in combination with white noise shocks. This new result has to be taken seriously, since it will be pointed out that the implied matching of the

¹ The estimation approaches of indirect inference (II) or the efficient method of moments (EMM) can be viewed as endogenizing this choice. On the other hand, this shifts the issue of arbitrariness, or judgement, to the choice of the auxiliary model that these methods employ. Carrasco and Florens (2002) provide a succinct overview of II, EMM and the method of (simulated) moments.

moments is entirely satisfactory.

The chapter is structured as follows. The next section introduces the MM estimation procedure together with a sketch of the bootstrap re-estimations that we will utilize. Section 3 describes the small New-Keynesian model to which this method is applied and lists the second moments the model is supposed to match. The results that we thus obtain are presented in Sections 4 and 5, where Section 4 deals with the period of the so-called Great Inflation and Section 5 with the Great Moderation. The main conceptual discussions are contained in Section 4, which is therefore subdivided into several subsections.

After contrasting the Bayesian with the MM estimation in Section 4.1, the next subsection examines in greater detail the problem of disentangling the endogenous and exogenous sources of inflation persistence. Section 4.3 subsequently employs a new econometric test by Hnatkovska et al. (2009) to decide whether our benchmark estimation is significantly superior to other, more special versions of the model. In Section 4.4 we temporarily step outside the model and ask if a still higher (composite) coefficient on lagged inflation would outperform the previous matching. Back in the original framework, Section 4.5 sets up the confidence intervals for the structural parameters, which invokes the abovementioned bootstrap re-estimations of the model because some of the parameters are estimated at their upper- or lower-bounds. In addition, this method allows us to compute a moment-specific p -value to characterize the model's validity. The organization of Section 5 for the Great Moderation period is similar, except that after the previous discussions the presentation of the results can now be much shorter. Section 6 concludes. Several more technical details are relegated to an appendix.

2.2 The moment matching estimation approach

As mentioned above, the MM estimation procedure computes a number of summary statistics, i.e. moments, for a model and searches for a set of parameter values that minimize a distance between them and their empirical counterparts. The method has also been applied to New-Keynesian DSGE models. The major part of this work is concerned with the matching of impulse-response functions (IRFs), where almost all of these contributions consider the responses to only one shock, namely, a monetary policy shock.² An exception is Altig et al. (2011), who add two types of technology shocks to the monetary impulse.

While this treatment avoids consigning itself to a choice about which other innovations to include in the modelling framework, a good matching of one type of IRFs does not necessarily imply a similar good match of another type. In this respect our situation will be different in that we deal with a model that has been subjected to a Bayesian estimation before. So the model has already as many shock

² Besides the early contribution by Rotemberg and Woodford (1987), examples from the last few years are Christiano et al. (2005), Boivin and Giannoni (2006), Henzel et al. (2009), Hülsewig et al. (2009). In contrast, Avouyi-Dovi and Matheron (2007) study the responses to a technology shock.

processes prespecified as there are endogenous variables. This allows us to consider a broader range of dynamic properties, which are conveniently summarized by the second moments of the economic key variables (which in the present case will be the output gap and the rates of interest and inflation). That is, we will be concerned with their unconditional contemporaneous and lagged auto-covariances and cross-covariances, which incidentally contain similar information to the IRFs of the (three) shock variables of the model.

Such a choice of moments has been more usual for the M(S)M estimation of, in a wider sense, real business cycle models (the ‘S’ refers to the cases where these moments cannot be computed analytically but must be simulated).³ Closest to our work is the MM estimation of a New-Keynesian model by Matheron and Poilly (2009). Their model is, however, richer than ours and instead of the output gap as a level variable they are interested in the comovements of the output growth rate. Hence one would have to be careful with a comparison of their results and ours.⁴

It may be emphasized that we fix our moments in advance and their number will not be too small, either. This commitment is different from an explicit moment selection procedure as it was, for example, used by Karamé et al. (2008). They begin with a large set of moments, estimate their model on them, and then step by step discard the moments which the model reproduces most poorly until an over-identification test fails to reject the model any longer.

Let us then turn to the moments that we adopt, which fortunately can be treated in an analytical manner. To explain this, we should first describe the general structure of our model. It is a hybrid variant of the New-Keynesian three-equations model, with forward-looking as well as backward-looking elements in the Phillips curve and the IS equation. Its canonical form reads,

$$\begin{aligned} A E_t y_{t+1} + B y_t + C y_{t-1} + v_t &= 0 \\ v_t &= N v_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon) \end{aligned} \tag{2.1}$$

The matrices A , B , C , N , Σ_ε with the structural parameters are here all ($n \times n$) square matrices (specifically, $n = 3$). The vector $y_t \in R^n$ contains the endogenous variables (with zero steady state values) and $v_t \in R^n$ collects the random shocks, which are supposed to be governed by an autoregressive process (certainly, N is a stable matrix). The i.i.d. innovations ε_t follow a normal distribution with a diagonal ($n \times n$) covariance matrix Σ_ε .

The equilibrium law of motion of (2.1) is described by the recursive equations

$$\begin{aligned} y_t &= \Omega y_{t-1} + \Phi v_t \\ v_t &= N v_{t-1} + \varepsilon_t \end{aligned} \tag{2.2}$$

³ These applications seem rather scattered, though; see Jonsson and Klein(1996), Hairault et al. (1997), Collard et al. (2002) and, more recently, Karamé et al. (2008), Gorodnichenko and Ng (2010), Ambler et al. (2011), Kim and Ruge-Murcia (2011).

⁴ Another difference is that they do not match directly the empirical second moments, which we do, but the moments deriving from the estimation of a canonical vector autoregression. This might somewhat favour a better match.

where Ω and Φ are two $(n \times n)$ matrices and Ω is required to be stable. Using the method of undetermined coefficients, Ω and Φ are successively obtained as the solutions to the following two matrix equations, which under determinacy are uniquely determined (I_n being the $(n \times n)$ identity matrix),

$$\begin{aligned} A\Omega^2 + B\Omega + C &= 0 \\ (A\Omega + B)\Phi + A\Phi N + I_n &= 0 \end{aligned}$$

As indicated, our aim in the moment matching estimation is that the stochastic process (2.2) reproduces the autocovariances of the empirical counterparts of the variables in the vector y_t . It is convenient in this respect that (2.2) is essentially a first-order vector autoregression (VAR). The theoretical autocovariances can thus be easily obtained from the closed-form expressions given, e.g., in Lütkepohl (2007). We only have to adjust the notation by changing the dating of the shocks and rewrite (2.2) as

$$\begin{bmatrix} y_t \\ v_{t+1} \end{bmatrix} = \begin{bmatrix} \Omega & \Phi \\ 0 & N \end{bmatrix} \begin{bmatrix} y_{t-1} \\ v_t \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} \varepsilon_{t+1} \quad (2.3)$$

With $z_t = (y_t', v_{t+1}')'$, $D = (0 \ I)'$, $u_t = D \varepsilon_{t+1}$, and A_1 the $(2n \times 2n)$ matrix on the right-hand side associated with the vector $(y_{t-1}', v_t')' = z_{t-1}$, eq. (2.3) can be more compactly written as

$$z_t = A_1 z_{t-1} + u_t, \quad u_t \sim N(0, \Sigma_u), \quad \Sigma_u = D \Sigma_\varepsilon D' \quad (2.4)$$

The (asymptotic) contemporaneous and lagged autocovariances of this VAR(1) are given by the matrices

$$\Gamma(h) := E(z_t z_{t-h}') \in R^{K \times K}, \quad K = 2n, \quad h = 0, 1, 2, \dots \quad (2.5)$$

Following Lütkepohl (2007, pp. 26f), their computation proceeds in two steps. First, $\Gamma(0)$ is obtained from the equation $\Gamma(0) = A_1 \Gamma(0) A_1' + \Sigma_u$, which yields

$$\text{vec } \Gamma(0) = (I_{K^2} - A_1 \otimes A_1)^{-1} \text{vec } \Sigma_u \quad (2.6)$$

(the symbol ' \otimes ' denotes the Kronecker product and invertibility is guaranteed since A_1 is clearly a stable matrix). Subsequently the Yule-Walker equations are employed, from which the lagged autocovariances are recursively obtained as

$$\Gamma(h) = A_1 \Gamma(h-1), \quad h = 1, 2, 3, \dots \quad (2.7)$$

The estimation seeks to match a subset of the coefficients in the matrices $\Gamma(h)$ to their observable empirical counterparts. In sum, let there be n_m of these moments, which are collected in a vector m . Furthermore, denote by θ the vector of the structural coefficients in (2.1) that are to be estimated, its dimension being n_θ . To make the dependence of the theoretical moments on the particular values of

θ explicit, we will write $m = m(\theta)$. On the other hand, let \widehat{m}_T designate the corresponding empirical moments from a sample of T observations. Below, reference will also be made to $\widehat{\Sigma}_m$ as an estimate of the covariance matrix of the moments (index T is here suppressed to ease notation).

The distance between the vectors of the model-generated and empirical moments is measured by a quadratic function that is characterized by an $(n_m \times n_m)$ weighting matrix W . Accordingly, the model is estimated by the set of parameters $\widehat{\theta}$ that minimize this distance over an admissible set $\Theta \subset R^{n_\theta}$, that is,⁵

$$\widehat{\theta} = \arg \min_{\theta \in \Theta} J(\theta; \widehat{m}_T, W) := \arg \min_{\theta \in \Theta} T [m(\theta) - \widehat{m}_T]' W [m(\theta) - \widehat{m}_T] \quad (2.8)$$

Regarding the weighting matrix in (2.8), an obvious since asymptotically optimal choice would be the inverse of an estimated moment covariance matrix (Newey and McFadden, 1994, pp.2164f). The optimality, however, does not necessarily carry over to small samples and a bias may arise in the estimations. As a consequence, in the context of estimating covariance structures even the identity matrix may be a superior weighting matrix (Altonji and Segal, 1996). In addition and not surprisingly in view of (2.7), with the choice of the above moments a matrix $\widehat{\Sigma}_m$ is so close to being singular that its inverse could not be relied on. The usual option in such a situation is to employ a diagonal weighting matrix the entries of which are given by the reciprocals of the variances of the single moments. This gives us

$$W_{ii} = 1 / \widehat{\Sigma}_{m,ii}, \quad i = 1, \dots, n_m \quad (2.9)$$

(and of course $W_{ij} = 0$ for $i \neq j$). Clearly, the less precisely a moment is estimated from the data, that is, the higher is its variance, the lower is the weight attached to it in the loss function. Since the width of the confidence intervals around the empirical moments $\widehat{m}_{T,i}$ is proportional to $(1/T)$ times the square root of $\widehat{\Sigma}_{m,ii}$, it may be stated that the model-generated moments $m_i(\widehat{\theta})$ obtained from the estimated parameters lie “as much as possible inside these confidence intervals” (Christiano et al., 2005, p. 17). Nevertheless, a formulation of this kind, which with almost the same words can also be found in several other applications, should not be interpreted too narrowly. In particular, it will be seen that a minimum of the loss function in (2.8) need not simultaneously minimize the number of moments outside the confidence intervals.

It is well-known that under standard regularity conditions the parameter estimates $\widehat{\theta}$ are consistent and asymptotically follow a normal distribution around the (pseudo-) true parameter vector θ^o . There is moreover an explicit formula in the literature (Newey and McFadden, 1994, pp.2153f) for estimates of the corresponding covariance matrix, which allows one to compute the standard errors of $\widehat{\theta}$ as the

⁵ The sample size T is included in the specification of the loss function to have the notation consistent with the literature that will be referred to below. It may also be added that if, in the course of the minimization search procedure for (2.8), some parameter leaves an admissible interval, it is reset to the boundary value, the distance of the thus resulting moments is computed, and then a sufficiently strong penalty is added that proportionately increases with the extent of the original violation. In this way also corner solutions to (2.8) can be safely identified.

square roots of its diagonal elements. In the present case, however, this approach faces two problems. First, it will turn out that locally the objective function J reacts only very weakly to the changes in some of the parameters. Hence these standard errors become extremely large and, beyond this (locally relevant) fact, are not very informative. The second point is that one of the regularity conditions will be violated if the minimizing parameter vector is a corner solution of (2.8); trivially, for some components i the distributions of the estimated parameters cannot be centred around the point estimates $\hat{\theta}_i$ then.

These reasons induce us to use a (parametric) bootstrap procedure as an alternative determination of standard errors or, more instructively, confidence intervals. To this end we work with the null hypothesis that the estimated model is the true data generating process. Thus, we take the estimated parameters $\hat{\theta}$ and, starting from the steady state (i.e. the zero vector), run a stochastic simulation of the model over $500 + T$ periods, from which the first 500 periods are discarded to rule out any transient effects. The underlying random number sequence may be identified by an integer index b . Repeating this a great number of times B , with different random number seeds of course, $b = 1, \dots, B$ artificial time series of length T are obtained. For each of them we compute the vector of the resulting moments, denoted as \hat{m}_T^b , and use their variances to set up the diagonal sample-specific weighting matrix W^b . Subsequently, for each b , the function $J(\theta; \hat{m}_T^b, W^b)$ is minimized over the parameter space Θ . Finally, the frequency distribution of the re-estimated parameters

$$\{ \hat{\theta}^b: b = 1, \dots, B \} \quad (2.10)$$

can serve as a proxy for the probability distribution of the $\hat{\theta}$. From (2.10), we can establish two types of 95% confidence intervals for the i -th component of the originally estimated vector $\hat{\theta}$, the standard percentile interval and Hall's percentile confidence interval. Hall's method has the advantage that it is asymptotically correct, but it may violate the admissible range of a parameter. Therefore we use Hall's interval if no such violation occurs and the standard interval otherwise. The details are spelled out in Appendix A2.

The bootstrap re-estimation experiment can also help us to decide at what significance level the null hypothesis may or may not be rejected. We only have to consider the frequency distribution of the values of the loss function,

$$J^b = J(\hat{\theta}^b; \hat{m}_T^b, W^b), \quad b = 1, \dots, B \quad (2.11)$$

and compare, let us say, the 95% quantile $J_{0.95}$ of (2.11) to the value $\hat{J} := J(\hat{\theta}; \hat{m}_T, W)$ that was obtained from the original estimation on the empirical moments in (2.8). At the conventional 5% significance level, the model would have to be rejected as being inconsistent with the data if \hat{J} exceeds $J_{0.95}$, otherwise it would have passed the test. In this way we can also readily construct a p -value of the model. It is given by the value of p that equates the $(1-p)$ -quantile of the distribution $\{J^b\}$ to \hat{J} , which says that if \hat{J} were employed as a benchmark for model rejection, then p is the error rate of falsely rejecting the null hypothesis that the model is true. Hence, in short, the higher this p -value the better the fit.

It goes without saying that these statements are conditional on the special choice of the moments that the model is required to match. Certainly, if more and more moments were added to our list, the p -value will dwindle.

2.3 The three-equations model

It should be explicitly made clear from the beginning that our estimations are concerned with a New-Keynesian model in gap form. That is, generally the trend rates π_t^* and r_t^* of inflation and interest (or the rates of these variables in a frictionless equilibrium) are allowed to vary over time, and what is showing up in the three key equations of the model are not the raw rates of inflation and interest π_t and r_t (i.e. their deviations from the zero steady state values in the simpler models), but the inflation gap $\hat{\pi}_t := \pi_t - \pi_t^*$ and the interest rate gap $\hat{r}_t := r_t - r_t^*$.⁶ There are several ways to interpret the occurrence of these more general gaps in, especially, the Phillips curve, and the persuasiveness of the microfoundations presently available for them in the literature is still another issue. We nevertheless join most of the empirical applications and leave this discussion aside. For simplicity, the trend variations themselves are treated as purely exogenous, so that π_t^* and r_t^* can remain in the background.

Regarding possible sources of persistence in the endogenous variables, which we then try to disentangle in the estimations, we concentrate on the Phillips curve. Here we include both lagged inflation in its deterministic core and serial correlation in the exogenous shocks. This is in contrast to the common practice that from the outset assumes either white noise shocks or purely forward-looking price setting behavior.⁷ On the other hand, the random shocks in the IS equation and the Taylor rule are supposed to be i.i.d. and persistence is only brought about by a lagged output gap and a lagged rate of interest, respectively. Denoting the output gap in period t by x_t , the model thus reads,

$$\begin{aligned}\hat{\pi}_t &= \frac{\beta}{1 + \alpha\beta} E_t \hat{\pi}_{t+1} + \frac{\alpha}{1 + \alpha\beta} \hat{\pi}_{t-1} + \kappa x_t + v_{\pi,t} \\ x_t &= \frac{1}{1 + \chi} E_t x_{t+1} + \frac{\chi}{1 + \chi} x_{t-1} - \tau (\hat{r}_t - E_t \hat{\pi}_{t+1}) + \varepsilon_{x,t} \\ \hat{r}_t &= \phi_r \hat{r}_{t-1} + (1 - \phi_r) (\phi_\pi \hat{\pi}_t + \phi_x x_t) + \varepsilon_{r,t} \\ v_{\pi,t} &= \rho_\pi v_{\pi,t-1} + \varepsilon_{\pi,t}\end{aligned}\quad (2.12)$$

The time unit is to be thought of as one quarter. The three shocks $\varepsilon_{z,t}$ are normally distributed around zero with variances σ_z^2 ($z = \pi, x, r$). All of the parameters

⁶ As for example remarked by Cogley et al. (2010, p. 43, fn 1) when discussing inflation persistence, it is not always completely plain in the literature whether the focus is on raw inflation or the inflation gap.

⁷ In similar models to ours, examples of excluding autocorrelated shocks in a hybrid Phillips curve are Lindé (2005), Cho and Moreno (2006) or Salemi (2006), while the purely forward-looking models studied by, e.g., Lubik and Schorfheide (2004), Del Negro and Schorfheide (2004), Schorfheide (2005) allow for some persistence in the shock process. We have chosen these references from the compilation in Schorfheide (2008, p. 421, Table 3).

are nonnegative. Specifically, β is the discount factor, κ a composite parameter that depends on the degree of price stickiness and assumptions on the production technology of firms, the coefficient α represents the degree of price indexation ($0 \leq \alpha \leq 1$), and the persistence in the supply shocks is given by the autocorrelation ρ_π ($0 \leq \rho_\pi < 1$).⁸ In the IS equation, χ is the representative household's degree of habit formation ($0 \leq \chi \leq 1$) and τ a composite parameter containing its intertemporal elasticity of substitution. In the Taylor rule, ϕ_r determines the degree of interest rate smoothing ($0 \leq \phi_r < 1$), and ϕ_x and ϕ_π are the policy coefficients that measure the central bank's reactions to contemporaneous output and inflation.

It depends on the particular kind of microfoundations whether or not α and χ also enter the determination of the composite parameters κ and τ , respectively, and whether the latter continue to be positive and well-defined in the polar cases $\alpha = 1$ or $\chi = 1$. In the estimations, however, κ and τ will not be subjected to any theoretical constraints in this respect.

The moments constituting the estimation of the model are based on the theoretical covariances of the interest rate gap \hat{r} , the output gap x and the inflation gap $\hat{\pi}$. Referring to the autocovariance matrices $\Gamma(h)$ from (2.6) and (2.7), we are thus concerned with the nine profiles of $\text{Cov}(p_t, q_{t-h}) = \Gamma_{ij}(h)$ for $p, q = \hat{r}, x, \hat{\pi}$ and, correspondingly, $i, j = 1, 2, 3$, while the lags extend from $h = 0, 1, \dots$ up to some maximal lag H . Given that the length of the business cycles in the US economy varies between (roughly) five and ten years, the estimations should not be based on too long a lag horizon. A reasonable compromise is a length of two years, so that we will work with $H = 8$. In this way we have a total of 78 moments to match: 9 profiles with $(1+8)$ lags, minus 3 moments to avoid double counting the zero lags in the cross relationships.

The empirical data on which the estimations of (2.12) are carried out derive from real GDP, the GDP price deflator, and the federal funds rate. To determine the exogenous trend rates underlying the model's gap formulation, we content ourselves with a deterministic setting and specify them by the convenient Hodrick-Prescott filter (as usual, although debatable, the smoothing parameter is $\lambda = 1600$).⁹

The total sample period covers the time from 1960 to 2007.¹⁰ Despite focussing on trend deviations instead of levels, one has to be aware that there are still great changes over these years in the variance of the three variables and partly also in

⁸ As it turns out, in some few estimations the fit could be improved by admitting negative values of ρ_π . We will, however, disregard this option since it seems too artificial, conceptually and since it implies a somewhat ragged profile of the autocovariances of the inflation rate.

⁹ Ireland (2007) and, more ambitiously, Cogley and Sbordone (2008) are two proposals of how to endogenize trend inflation as the target set by the central bank. Ireland (p. 1864), however, concludes from his estimations that still "considerable uncertainty remains about the true source of movements in the Federal Reserve's inflation target". Laubach and Williams (2003) and Messonnier and Renne (2007) are attempts at an estimation of a time-varying natural rate of interest.

¹⁰ The Hodrick-Prescott trend is computed over a longer period, to avoid end-of-period effects. The time series of the gaps that we thus obtain can be downloaded from http://www.bwl.uni-kiel.de/gwif/downloads_papers.php?lang=en (if this string is copied into the browser address bar, the underscore character '_' may have to be retyped manually).

the pattern of their cross covariances. This makes it necessary to subdivide the period into two subsamples, which are commonly referred to as the periods of the Great Inflation (GI) and the Great Moderation (GM). We define the former by the interval 1960:1–1979:2 and the latter by 1982:4–2007:2; the time inbetween is excluded because of its idiosyncrasy (Bernanke and Mihov, 1998). To give an immediate example for the need of the subdivision, the standard deviation of the annualized inflation gap in GI is 1.41% *versus* 0.77% in GM; for the output gap it is 1.77% in GI *versus* 1.15% in GM.

2.4 The Great Inflation period

2.4.1 Basic results

The three-equations model (2.12) includes 12 structural parameters. Among them, the discount factor β is not a very critical coefficient and is therefore directly calibrated at $\equiv 0.99$. So the following 11 parameters remain to be estimated: α , κ , ρ_π , σ_π in the Phillips curve and its shock process; χ , τ , σ_x in the IS equation; and ϕ_π , ϕ_x , ϕ_r , σ_r in the Taylor rule. The inflation and interest rate gap in (2.12) are annualized, which may be taken into account when considering the order of magnitude of κ , τ , ϕ_x and the two noise levels σ_π , σ_r .

We begin with a Bayesian reference estimation (BR) of the model. The mean values of the posterior distribution of the parameters are reported in the first column of Table 2.1 (the priors are documented in Appendix A1). Except perhaps for the relatively high policy coefficient ϕ_x , the results are not dramatically different from other Bayesian estimations in the literature. In particular, regarding the sources of inflation persistence, low coefficients on expected inflation in the Phillips curve (i.e., low values of α) and a high autocorrelation ρ_π in the shock process are typical for them.¹¹ It is, however, interesting to note an exception to this rule. Del Negro et al. (2007, p.132, Table 1) obtain high price indexation ($\alpha = 0.76$) and low shock persistence ($\rho_\pi = 0.12$), despite their setting of rather opposite priors.¹² This outcome exemplifies that even within the Bayesian framework, the tendency towards a purely forward-looking Phillips curve with persistent random shocks is possibly not an unequivocally established property, yet.

The original motivation of this chapter was to check the role of α and ρ_π from the outside, by an alternative estimation approach. The pivotal result of our MM estimation is given in the second column of Table 2.1, which we will refer to as estimation A, or model A. As a matter of fact, the most immediate observation is on α and ρ_π , for which the contrast to the Bayesian estimation could not be more striking: α is estimated at its maximum value of unity and ρ_π at its minimum value of zero.

¹¹ For examples from more general models, see Smets and Wouters (2003, 2007), Adolfson et al. (2007), Benati and Surico (2007), Fève et al. (2009), Cogley et al. (2010). Apart from the determination of trend inflation, estimation BR can be directly compared to Castelnuovo's (2010) results for his so-called TI model, on which he (arguably) imposes $\alpha = 0$.

¹² The present symbols α and ρ_π correspond to their ι_p and ρ_{λ_f} .

Table 2.1: Parameter estimates for GI.

	Bayesian	Moment Matching		
α	0.067 (0.000 – 0.144)	1.000 (0.543 – 1.000)	0.000	0.700
κ	0.198 (0.119 – 0.275)	0.051 (0.000 – 0.085)	0.242	0.067
ρ_π	0.552 (0.433 – 0.668)	0.000 (0.000 – 0.334)	0.692	0.550
σ_π	0.666 (0.492 – 0.827)	0.571 (0.242 – 0.838)	0.664	0.274
χ	0.758 (0.672 – 0.844)	1.000 (0.730 – 1.000)	1.000	1.000
τ	0.034 (0.018 – 0.049)	0.096 (0.008 – 0.155)	0.062	0.080
σ_x	0.612 (0.511 – 0.706)	0.610 (0.296 – 0.887)	0.440	0.590
ϕ_π	1.173 (1.000 – 1.361)	1.482 (1.338 – 1.690)	1.524	1.574
ϕ_x	1.336 (0.705 – 1.965)	0.030 (0.000 – 0.210)	0.000	0.068
ϕ_r	0.792 (0.717 – 0.866)	0.333 (0.136 – 0.457)	0.421	0.383
σ_r	0.729 (0.630 – 0.827)	0.000 (0.000 – 0.595)	0.000	0.000
J	209.1	47.6	119.4	77.6
MCI missed	21	0	5	1

Note: The discount factor is $\beta = 0.99$ throughout. In estimations B and C, α is fixed at 0.00 and 0.70, respectively. The smaller numbers indicate the confidence intervals; from the posterior distribution in a Bayesian reference estimation (BR), while in estimation A they are computed from (A1) in Appendix A2 for α , ρ_π , χ , ϕ_x , σ_r and from (A2) for κ , σ_π , τ , σ_x , ϕ_π , ϕ_r . The last row gives the number of moments ('M') that miss the confidence intervals ('CI') of the empirical moments. The bold face figures emphasize certain results (model A) or assumptions (model B and C).

Before we turn to a more comprehensive discussion of these parameters and the other results in the table, let us consider the matching properties of estimations BR and A. While it is trivial that BR implies a higher loss J than model A, the differences are so substantial that in effect the two estimation approaches may appear to concentrate on rather distinct features of the data, which show no general tendency to imply each other. This is, however, a preliminary and informal evaluation. In Section 2.4.3 a rigorous econometric test will be applied in order to see whether or in what sense it can be maintained.

It is one purpose of Figure 2.1 to illustrate the differences from the point of view of moment matching. The thin dashed lines in the diagrams are the empirical auto- and cross-covariances of the interest rate, output and inflation (since there will be no more risk of confusion, we will from now on omit the expression ‘gap’ when discussing these variables). The shaded area is the 95% confidence band around them. The bold (red) lines depict the moments obtained from the MM estimation A, while the dotted (blue) lines are the moments implied by the Bayesian estimation BR. Recall that in order to evaluate their goodness-of-fit as our loss function defines it, only the first eight lags are relevant.

Inspecting the performance of the MM estimation with the naked eye, the match it achieves looks very good over the first few lags and still fairly good over the higher lags until the maximal lag $H = 8$. In any case, it is remarkable that all of the moments are contained within the confidence intervals of the empirical moments. This even holds true for the covariances up to lag 20. Hence, at the usual 5% significance level and as far as the (asymptotic) second moments are concerned that we chose, the model could not be rejected as being inconsistent with the real-world data generation process.

In finer detail, the model-implied moments show less persistence than the empirical covariances, in that they return more quickly to the zero level and then stay there. In other words, with respect to the covariances of its state variables the model predicts a shorter memory than it seems to prevail in reality. Reproducing a longer memory would, however, ask too much from a small model such as the present one, if the longer memory is a reliable phenomenon at all.

The covariances implied by the parameters of the Bayesian estimation are far less satisfactory. In sum, as reported in the first column of Table 2.1, 15 of their moments are outside the empirical confidence intervals, although the violation is not overly strong.¹³ The best match, actually a very good one, is obtained for the auto-covariances of inflation, $\text{Cov}(\hat{\pi}_t, \hat{\pi}_{t-h})$. Still acceptable is the persistence in these statistics for output and the interest rate, while their initial levels are too low. Mainly responsible for the high value of the loss function ($J = 209.1$) in Table 2.1 are the cross-covariances, the performance of which is rather poor, especially if one has a look at the practically vanishing $\text{Cov}(x_t, \hat{\pi}_{t\pm h})$ statistics. Conclusions from the Bayesian estimation that concern the central features of the dynamic output-inflation nexus may therefore be taken with some care; at least in

¹³ The highest t -statistic is around 2.30.

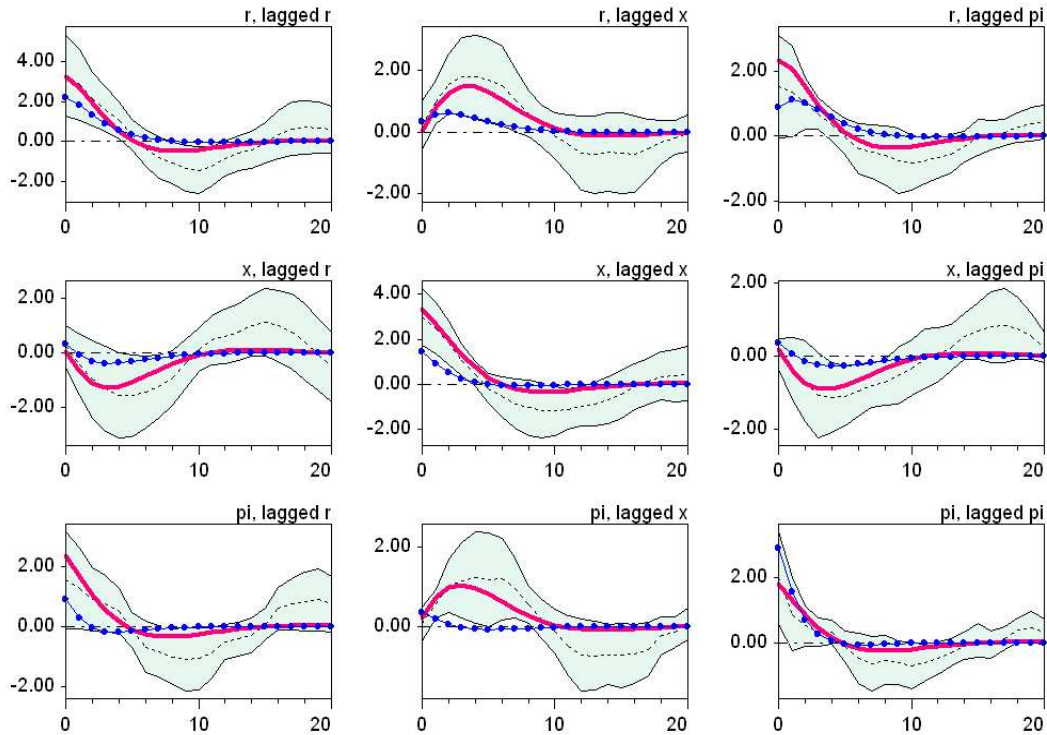


Figure 2.1: Estimated *versus* empirical covariance profiles (GI).

Note: The bold (red) line results from the MM estimation A of Table 2.1, the solid (blue) line with dots from the Bayesian reference estimation BR. The shaded area is the 95% confidence band around the empirical moments.

the present context the relatively good one-period ahead forecasting properties of this approach do not seem well suited to deliver authoritative statements about the general interrelationships of these variables.¹⁴

2.4.2 Price indexation *versus* shock persistence

The MM estimation makes a definite statement about the relative importance of price indexation and the shock autocorrelation as the two main sources of persistence in the Phillips curve. The outcome of $\alpha = 1$ and $\rho_\pi = 0$ is the exact opposite of the message from the papers by, for example, Ireland (2007, p.1864) and Cogley and Sbordone (2008, p.2113), who found no significant evidence for backward-looking behavior in similar price setting specifications. They argue that

¹⁴ With respect to likelihood methods in general, the different properties of estimation A and BR tend to contradict the intuition expressed, for example, by Schorfheide (2008, p.402) that “[s]uperficially, the likelihood function peaks at parameter values for which a weighted discrepancy between DSGE model-implied autocovariances of [state vector] x_t and sample autocovariances is minimized.”

a purely forward-looking Phillips curve proves fully sufficient because their models appropriately account for time-variation in the inflation target, which can substitute for the backward-looking terms in previous estimations on raw inflation data or their deviations from the mean.

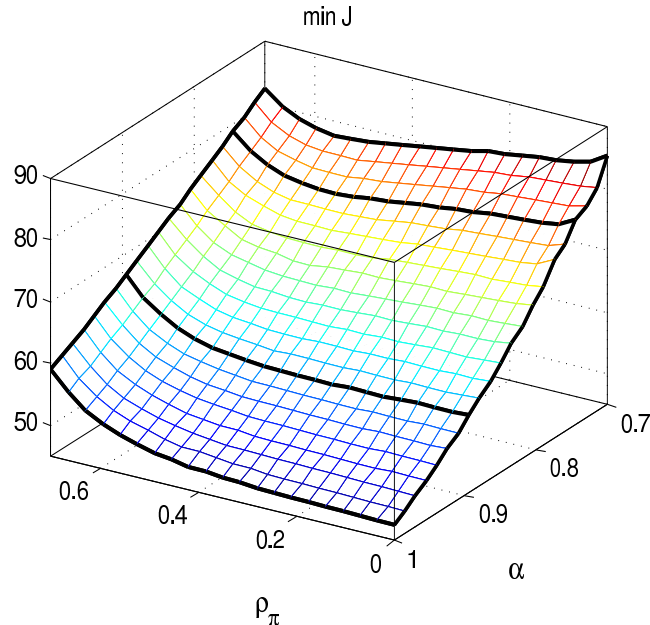


Figure 2.2: Minimized values of J given α and ρ_π (GI).

Since our inflation gap variable is based on a time-varying trend, too, the contradistinctive results appear somewhat puzzling. There are several possible explanations for this, beginning with different estimation methods and different sample periods.¹⁵ Also the specific details in the Phillips curves may be less innocent than a short description of their basic ingredients suggests. Another point makes things even more complicated, which is to realize that identification of forward- and backward-looking terms in a Phillips curve may easily depend on assumptions about other structural equations in a general equilibrium model, including the precise auxiliary assumptions about the shock processes. To paraphrase the concluding sentence in Beyer and Farmer (2007, p. 527), any attempt to categorize an observed data series as arising from two different Phillips curve specifications “is determined as much by subtle choices over the way to model the dynamics as it is by the data themselves”.¹⁶ Our estimation is therefore far from being able to settle the controversial subject of backward-looking *versus* forward-looking behavior. For the time

¹⁵ In particular, Ireland and Cogley & Sbordone estimate their models over longer sample periods, namely 1959:1–2004:2 and 1960:1–2003:4, respectively. The common wisdom is that for the years after 1984, the New-Keynesian Phillips curve needs to explain only a moderate degree of persistence. We may, however, anticipate that in our estimations of the Great Moderation below the coefficient on lagged inflation in the Phillips curve is not driven to zero, either.

¹⁶ Their paper illustrates this with the distinction between determinacy and indeterminacy.

being, we can only point out the strikingly different results and must leave it to further effort to find out more about what essentially is responsible for them.

Within the present framework, one may now scrutinize the reliability of the estimates $\alpha = 1$ and $\rho_\pi = 0$. Because of their common role to generate persistence in the Phillips curve, the two parameters are also the first candidates the variations of which might give rise to multiple local minima. This idea motivates the following complementary estimations: treat both α and ρ_π as exogenous parameters, consider a grid of the pairs (α, ρ_π) , and estimate the nine remaining parameters for each of the grid points.

Figure 2.2 plots the thus minimized values of J in the three-dimensional space above the (α, ρ_π) -plane, for $0.70 \leq \alpha \leq 1.00$ and $0.00 \leq \rho_\pi \leq 0.70$. What immediately leaps to the eye is the perfect smoothness of the surface and the absence of any local valley. Overall, Figure 2.2 can instil additional confidence in us that the corner point $(\alpha, \rho_\pi) = (1.00, 0.00)$ does indeed constitute the global minimum.

A second feature of Figure 2.2 can shed more light on the informal question for the relative importance of price indexation (α) *versus* the persistence in the shock process to inflation (ρ_π). In the present context, ‘importance’ may be measured by the relative changes in $\min J$ brought about by the variations in α and ρ_π . The bold lines on the surface along the ρ_π -axis clearly show that, for fixed values of α , the variations in ρ_π have only a minor impact on the goodness-of-fit, at least for values of ρ_π in the range between 0.00 and 0.40, say. For fixed values of ρ_π , on the other hand, the deterioration is much more serious when α is gradually decreased. Indexation is therefore a crucial parameter for the moment matching and higher persistence in the shocks is not nearly capable of making up for the negative effects of lower indexation. As this is a global phenomenon in GI, the best fit for this period entails maximal price indexation, $\alpha = 1$.

After establishing indexation as the parameter of primary concern in the Phillips curve, it is interesting to see the changes in the estimation results when only α is exogenously varied and J is minimized across the remaining ten parameters, which now include ρ_π . Figure 2.3 presents the most important reactions. First of all, the loss function in the upper-left panel is monotonically rising as α decreases over the entire admissible range from unity down to zero. This underlines what has just been said about the dominance of the effects from α over the effects from ρ_π , not only locally but over the full domain of α . The worsening from $J = 47.6$ at $\alpha = 1$ to $J = 119.4$ at zero indexation (cf. estimation B in Table 2.1) appears rather severe, though a discussion of whether it can also be categorized as statistically significant will be postponed until the next subsection.

The next effect of interest are the implied changes in the autocorrelation ρ_π of the shocks. As expected, lower indexation gives more scope for higher shock persistence, and again this holds over the entire range of α ; see the upper-right panel in Figure 2.3. It is, however, remarkable that between $\alpha = 0.95$ and $\alpha = 0.94$ an almost discontinuous change in the optimal value of ρ_π occurs, when ρ_π jumps from 0.051 to 0.262. The reason for this is that the functions $\rho_\pi \mapsto J(\rho_\pi)$ in Figure 2.2 for fixed

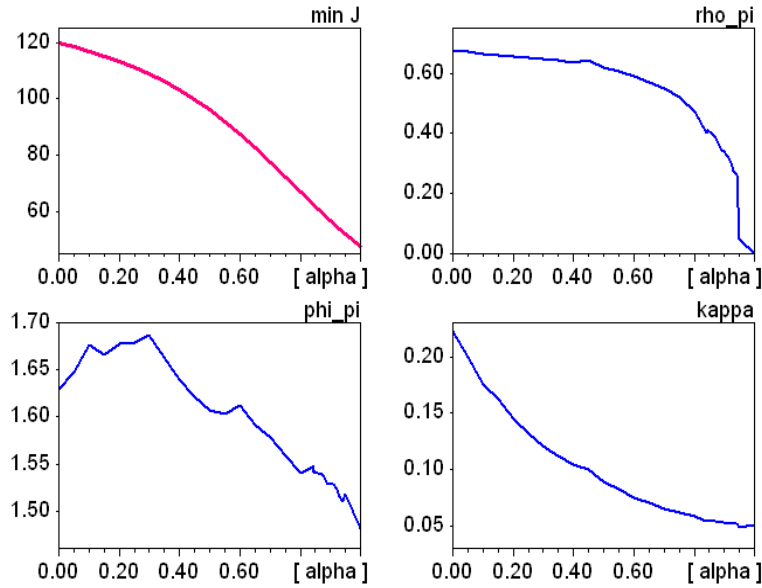


Figure 2.3: MM estimation of the model under exogenous variations of α (GI).

values of α are all very flat in that region, which implies that already small changes in their shape brought about by small changes in α can shift the minimum of these functions considerably.¹⁷

Our reasoning concerning the Phillips curve has so far left aside the output gap as a source of inherited persistence. The lower-right panel in Figure 2.3 for the optimal values of the parameter κ reveals a stronger influence of this variable as compensation for a reduced persistence from price indexation.

The results illustrated in these three panels can be related to Fuhrer’s (2006) analysis of the constituent factors contributing to inflation persistence. For this, he concentrates on the autocorrelations of the inflation rate as they are brought about by a hybrid Phillips curve and a simple AR(1) process for the driving variable. Our study is more general in that it incorporates additional criteria the model is desired to match, and also discusses the possible influence of persistence in the shock process to inflation.¹⁸ Fuhrer’s main message from his GMM and maximum likelihood estimations is nevertheless maintained: little is inherited from the persistence of (the shock and) the driving variable—and if so, this deteriorates the performance of the model. Hence, “the predominant source of inflation persistence in the NKPC is the lagged inflation term” (Fuhrer, 2006, p. 79). Actually, his coefficient on lagged inflation is typically even higher than 0.5025, which is the maximal value that we can get in eq. (2.12) when $\alpha = 1$. This is a numerical issue that we return to in Section 2.4.4.

¹⁷ It actually required special care to spot the jump of the optimal ρ_π precisely between 0.94 and 0.95.

¹⁸ Fuhrer assumes white-noise i.i.d. shocks and makes a remark that the serial correlation that might be added to the shock variable will plausibly be relatively low (Fuhrer, 2006, p. 70).

Among the other parameters in the estimations of the model and their reactions to diminished indexation, the lower-left panel of Figure 2.3 shows the policy coefficient ϕ_π on the inflation gap in the Taylor rule. Higher values of it might be interpreted as an indirect source of inflation persistence, acting through the interest rate channel. This point of view is confirmed by the moderate increase of ϕ_π in response to a reduction in α . Nevertheless, as indexation decreases further, other mechanisms become more influential and eventually reverse this effect. Besides, the estimated order of magnitude of ϕ_π (and also ϕ_x) appears to be more reasonable for MM than BR.

2.4.3 Is full price indexation significantly superior?

In the discussion of Figure 2.1 we have emphasized the much better match of our estimation A with price indexation $\alpha = 1$ *versus* the Bayesian reference estimation BR with an indexation close to zero. In terms of the loss function, this amounts to a comparison of $J = 47.6$ *versus* $J = 209.1$. In the previous subsection, when assessing the role of α in finer detail, it has furthermore been pointed out that imposing the purely forward-looking case $\alpha = 0$ on the MM estimation deteriorates J from 47.6 to 119.4 (see Table 2.1). Nevertheless, these figures as such are not yet sufficient to characterize the differences as ‘significant’. Especially because J is a quadratic function of the moment deviations, the apparently large differences might be somewhat misleading.

Table 2.1 also reports that the two models BR and B have, respectively, 21 and 5 of the model-generated moments outside the empirical confidence intervals. Since all of the moments of model A are inside the intervals, it might be said that this model cannot be strictly told apart from the hypothetical true data generation process, whereas the matching obtained for models BR and B can. On the other hand, this need not necessarily imply that BR and B are significantly inferior to the unconstrained model. For example, we would hesitate to subscribe to this statement if, in the comparison of two models, the set of critical moments were close to the boundaries of the confidence intervals—one inside, the other outside the intervals.

As a matter of fact, as has been remarked above (see footnote 13), the violations of the confidence interval conditions by model BR are not very strong, and a similar statement holds true for model B. In order to decide whether these estimations are significantly inferior to model A, a test procedure for MM-estimated models proposed by Hnatkovska, Marmer and Tang (2009; HMT henceforth) seems tailor-made for the present framework; although the comparison of model A and BR requires a slight modification of the latter, which is explained further below. It is particularly charming that the authors are explicitly concerned with misspecified models.¹⁹

¹⁹ See Definition 2.1 in HMT for a precise definition of misspecification, which is here moment-specific. There is no reason to believe that a small macroeconomic model such as (2.12) should not satisfy it, despite the conventional formulation above that model A “cannot be rejected by the data”.

The following description recapitulates what is needed to apply the econometric theorems of HMT as a recipe. To set the stage in general, let X and Y be two arbitrary models that are estimated on the same set of empirical moments. With respect to $I = X, Y$, let θ^I be the vector of free parameters entering model I and $m^I(\theta^I)$ the vector of the moments generated by θ^I in model I . Three cases need to be distinguished: (a) Model Y is nested in model X , which means that for all moments $m^Y(\theta^Y)$ there is a parameter vector θ^X with $m^X(\theta^X) = m^Y(\theta^Y)$; (b) X and Y are strictly non-nested, which means they have no moment vector in common; (c) X and Y are overlapping, according to which the models are non-nested and have at least one moment vector in common.

As our estimations were laid out, model A nests model B with its constraint $\alpha = 0$. Model A 's optimal value of α is, however, a corner solution ($\hat{\alpha} = 1$), whereas the test statistics put forward by HMT assume that the estimated parameters are in the interior of the admissible region (see their Assumption 2.5(b)). Hence α must be treated as being exogenously fixed at unity, by which the two models become strictly non-nested or overlapping. The same applies to any model and to any of its parameters that has been estimated at an end-point of the admissible interval.

The basic question of the model comparison approach is whether the lower value of the loss function of a model indicates a significantly superior performance. If the models are nested or overlapping, an answer first has to carry out a test that establishes whether or not model X and Y have the same pseudo-true moments. If they have, one concludes that the two models have the same fit and the testing is done. If not, and the models are nested, unequal moment vectors also mean rejection of the null hypothesis of an equal fit; that is, under these circumstances the model with the lower loss has a significantly superior fit.

On the other hand, if the moments are found to be significantly different in the overlapping case, or if they are strictly non-nested, the fit of one model might still be similarly good (or bad) to the fit of the other model. It is now the task of another step to decide on the significance of the difference in the loss.

Both steps in the test procedure are based on a direct comparison of the loss functions of the two models, which in the present context we write as

$$J^I(\theta^I; \hat{m}_T, W) := T [m^I(\theta^I) - \hat{m}_T]' W [m^I(\theta^I) - \hat{m}_T], \quad I = X, Y \quad (2.13)$$

(recall that \hat{m}_T is the vector of the empirical moments). Letting X be the candidate of a significant superiority, reference is made to the (positive and scaled) difference between the two minimized values of J^Y and J^X , as they are brought about by $\hat{\theta}^Y$ and $\hat{\theta}^X$, respectively. HMT use the acronym QLR for it (alluding to the term 'quasi-likelihood ratio').²⁰ With respect to the notation in (2.13), it is defined as

$$\text{QLR}(\hat{\theta}^Y, \hat{\theta}^X) := (1/T) [J^Y(\hat{\theta}^Y; \hat{m}_T, W) - J^X(\hat{\theta}^X; \hat{m}_T, W)] \quad (2.14)$$

In the first step, for two nested or overlapping models, HMT derive an explicit expression for the probability distribution P to which $T \cdot \text{QLR}$ converges in probability

²⁰ Since we only use QLR as a recipe, notational reference to the sample length T , which is helpful for the formulation of asymptotic statements, is suppressed.

under the null hypothesis that both models have the same pseudo-true moments (formally introduced shortly below). This proposition involves the weighting matrix W , the covariance matrix of the moments Σ_m and its square root $\Sigma_m^{1/2}$, and two special and rather complicated matrices V^X, V^Y pertaining to model X and Y, respectively; all of these matrices have format $(n_m \times n_m)$.²¹ The probability element is represented by a random vector z the n_m components of which are independent and follow the standard normal. Then, the asymptotic distribution P we are looking for reads,

$$P \sim z' \Sigma_m^{1/2} W (V^Y - V^X) W \Sigma_m^{1/2} z \quad (2.15)$$

In their paper, HMT describe P as a mixed χ^2 distribution. The term is somewhat delusive as the latter has a positive support, while from P also negative values could be obtained with positive probability, even if model Y is nested in model X.²² Intuitively, this may happen if, compared to the estimated model Y, the estimate $\hat{\theta}^X$ of model X does not lead to an equal or superior match in all of the moments. In this case a non-negligible subset of the realizations of the vector z can put sufficient weight on exactly the moments in which model X is slightly inferior to Y.

The distribution P is nonstandard and, in particular, depends on the unknown true moments and their covariance matrix. However, the distribution and its critical values can be approximated by simulations that use (a) consistent estimates of the matrices entering P , and (b) sufficiently many random draws of the vector z . To be more explicit, let a hat over Σ_m, V^X, V^Y denote the estimates of these matrices (Appendix A3 and A4 give the further details), and consider $c = 1, \dots, 1000$ random draws $z_c \in R^{n_m}$ from the multivariate standard normal. This gives us a collection of 1000 realizations of the estimated version of (2.15),

$$\{ z_c' \hat{\Sigma}_m^{1/2} W (\hat{V}^Y - \hat{V}^X) W \hat{\Sigma}_m^{1/2} z_c : z_c \sim N(0, I_{n_m}), c = 1, \dots, 1000 \} \quad (2.16)$$

It is the 95% quantile of these simulated values, which may be designated $Q_{0.95}$, that enables us to test whether the two models have identical pseudo-true moments, that is, whether the hypothesis

$$m^Y(\theta^{Y,o}) = m^X(\theta^{X,o}) \quad (2.17)$$

is satisfied, where $\theta^{I,o}$ are the pseudo-true parameters of model I ($I = X, Y$).²³ Accordingly, at a 5% significance level, the recipe is:

$$\text{reject (2.17) if } T \cdot \text{QLR}(\hat{\theta}^Y, \hat{\theta}^X) > Q_{0.95} \quad (2.18)$$

²¹ To be precise, for the following HMT suppose that the covariance matrix Σ is positive-definite, whereas our moments are not independent so that our estimated $\hat{\Sigma}$ is only semipositive-definite (which we checked). HMT employ the assumption to ensure that the test statistics involving $\hat{\Sigma}$ are strictly positive (private communication with Vadim Marmer). However, the condition is by no means necessary for that. If Σ is semipositive-definite then, for reasons of continuity, all of the statements remain true if the statistics happen to be nonzero, the only possible difference being that a strict inequality may turn into a weak inequality.

²² Vadim Marmer clarified this point to us in a private communication, where he also identified this possible phenomenon in a formal decomposition of the QLR statistic.

²³ Formally, with respect to the notation in eq. (2.8) and to m^o as the moment vector resulting from the unknown true model of the economy, $\theta^{I,o}$ satisfies $J^I(\theta^{I,o}; m^o, W) \leq J^I(\theta^I; m^o, W)$ for all θ^I in the set of feasible parameters.

If (2.18) applies and the models are nested then, as mentioned above, we can at the same time conclude that the one with the lower loss succeeds in a significantly better fit. On the other hand, a failure to reject (2.17) tells us that the two models have essentially the same fit, so that the testing is completed.

Taking the second step in the test procedure, let us suppose that the inequality in (2.18) is satisfied, or that we already know that model X and Y are strictly non-nested. Regarding the relative quality of the fit, the precise formulation of the null and the alternative hypothesis, H_0 and H_A , reads,

$$\begin{aligned} H_0 &: J^X(\theta^{X,o}; m^o, W) = J^Y(\theta^{Y,o}; m^o, W) \\ H_A &: J^X(\theta^{X,o}; m^o, W) < J^Y(\theta^{Y,o}; m^o, W) \end{aligned} \quad (2.19)$$

where m^o is the vector of the moments generated by the unknown true data generation process of the economy. The test of (2.19) utilizes QLR once again to set up a t -statistic. To this end, the following estimate of an asymptotic standard deviation is specified,

$$\hat{s} = 2 \cdot \sqrt{\{\hat{\Sigma}_m^{1/2} W[m^Y(\hat{\theta}^Y) - m^X(\hat{\theta}^X)]\}' \{\hat{\Sigma}_m^{1/2} W[m^Y(\hat{\theta}^Y) - m^X(\hat{\theta}^X)]\}} \quad (2.20)$$

Letting $z_{1-0.05/2}$ be the conventional critical quantile of the standard normal distribution, the second step of the model comparison procedure is:

$$\text{reject } H_0 \text{ in favour of } H_A \text{ if } \sqrt{T} \text{QLR}(\hat{\theta}^Y, \hat{\theta}^X) / \hat{s} > z_{1-0.05/2} = 1.96 \quad (2.21)$$

To sum up, HMT's model comparison test is constituted by the results from (2.18) and, if the second step is still to be taken, from (2.21).

When now, in a first application, we want to compare our model A to the Bayesian reference model BR, we meet with the obstacle that BR has not been estimated by MM. To fit BR into the MM framework, we help ourselves by fixing all of the numerical parameters of BR except σ_π , which is treated as the one and only free parameter for an MM estimation. The value that thus minimizes the loss function changes slightly from 0.666 to $\sigma_\pi = 0.690$, reducing the loss from 209.1 to 208.3. Let us call this modified model BR',

$$\sigma_\pi = 0.690, \quad \text{other parameters from BR} \quad (\text{BR}')$$

and instead of BR, compare model A to BR'.²⁴ Clearly, A and BR' are non-nested, though we do not know whether they are strictly non-nested or overlapping. Since the latter cannot be ruled out, we should begin with computing the statistics needed for the test in eq. (2.18). The basic figures are reported in the first two rows of Table 2.2. First, the difference between the minimized values of J , which equals $T \cdot \text{QLR}$, clearly exceeds the 95% quantile $Q_{0.95}$ of the simulated test distribution (2.16). At the 5% significance level we can therefore discard the hypothesis that model A and BR' have equal moments in the sense of eq. (2.17), so that we continue with step 2 of the test.

²⁴ For model A, the parameters α , ρ_π , χ , ϕ_x , σ_r are exogenously fixed since they were estimated at (or close to) the boundary of their feasible range.

For the standard deviation in (2.20), $\hat{s} = 13.09$ is obtained. Together with $\sqrt{T} \text{QLR} = T \cdot \text{QLR} / \sqrt{T} = 160.7 / \sqrt{78} = 18.20$, the test statistic in (2.21) is computed as 1.39. As this falls short of the critical value, we are not legitimated to conclude that the moment matching implied by the slightly modified Bayesian estimation BR' with $J = 208.3$ is significantly inferior to the match of our basic MM estimation A with $J = 47.6$, even though the two models are sure to have different moments. The same result is obtained when comparing model A with the MM estimation B of the purely-forward-looking model variant, which has $\alpha = 0$ imposed.²⁵

Table 2.2: Comparison of alternative estimations.

Model	α	J	$T \cdot \text{QLR} $	$Q_{0.95}$	$\sqrt{T} \text{QLR} / \hat{s}$	Conclusion
<u>GI:</u>						
A	1.00	47.6	--	--	--	--
BR' vs. A	0.07	208.3	160.7	130.1	1.39	different moments, but equivalent fit
B vs. A	0.00	119.4	71.8	39.0	1.51	different moments, but equivalent fit
C vs. A	0.70	77.6	30.0	29.8	--	same moments (at the 5% margin)
F' vs. BR'	2.48	13.0	195.3	149.3	--	different moments, F' superior to BR'
F vs. B	2.48	12.7	106.7	48.3	--	different moments, F superior to B
F vs. A	2.48	12.7	34.9	21.8	--	different moments, F superior to A
<u>GM:</u>						
A	0.82	54.1	--	--	--	--
BR' vs. A	0.03	157.7	103.6	121.7	--	same moments
B vs. A	0.00	68.4	14.3	50.6	--	same moments

Note: Models F, F' for GI and A, B for GM are introduced below. Column α reproduces the values for the first model.

An intuitive argument to understand this finding is that there are some moments of the two models that are on opposite sides of the profile of the empirical moments. This holds for a comparison of A and BR' as well as A and B. So the moments are relatively far apart from each other, while their deviations from the empirical moments are more moderate. The first phenomenon contributes to the

²⁵ Ireland (2007, p. 1864) with his maximum likelihood approach obtains a significant result to the opposite. As already indicated above, in his estimations the parameter α leans up against its lower bound of zero. He checked this estimate by alternatively imposing the constraint $\alpha = 1$ and found that this specification was firmly rejected by a likelihood ratio test.

overall conclusion of significantly distinct moments of, say, model A and B in the first step of the test procedure. The latter deviations are evaluated by the loss function as $J^A(\hat{\theta}^A; \hat{m}_T, W)$ and $J^B(\hat{\theta}^B; \hat{m}_T, W)$, respectively, and although naively the difference between these two values may appear rather large, the second step of eqs (2.20), (2.21) does not yet classify it as significant. If this is not exactly what one has expected then, given the empirical and asymptotic moments of the two estimations, the failure of the inequality in (2.21) to hold true might be viewed as being due to the fact that our sample size $T=78$ is too small.²⁶

After establishing that the two MM estimations A and B yield at least significantly different moments, let us utilize once more the first step of the model comparison test. Again treating the degree of price indexation α as an exogenous parameter, we gradually increase it from $\alpha = 0$ and ask from what value of α on do the moments from the corresponding estimations differ no longer significantly from the moments of model A with $\alpha = 1$. The borderline case is brought about by $\alpha = 0.70$, which gives rise to estimation C in Table 2.1. As shown in Table 2.2, the resulting test statistic $T \cdot \text{QLR}(\hat{\theta}^C, \hat{\theta}^A)$ is 30.0 and thus essentially equal to the 95% quantile $Q_{0.95} = 29.8$ of the simulated distribution from (2.16). Estimations where α is fixed at higher values than 0.70—and only these—lead to $T \cdot \text{QLR} < Q_{0.95}$ and therefore do not reject the hypothesis of equal moments.

The basic feature of these model comparisons is the scope for obtaining significantly different moments, which was established in the first step of the test procedure for overlapping models. The second step, however, showed that this is not yet sufficient to conclude that the model with the lower loss is also significantly better than the other. Hence, if we like to get the more pronounced result of one model significantly outperforming the other, we have to broaden the framework of the discussion. This is an issue that we can return to below.

2.4.4 Admitting stronger backward-looking behavior

Having identified the momentous role of full indexation in the price adjustments of the non-optimizing firms, we may take one step further. In fact, the unchecked fall of the function $\alpha \mapsto \min J$ towards the end-point $\alpha = 1$ in the top-left panel of Figure 2.3 suggests that still higher values of α would lead to a further improvement in the matching of the moments. This idea could be pursued in another framework that allows for wider intervals of the two coefficients on expected and lagged inflation in the Phillips curve. In the simplest case, a parameter $\mu \in [0, 1]$ may be introduced and the coefficients on $E_t \hat{\pi}_{t+1}$ and $\hat{\pi}_{t-1}$ directly specified as $(-\mu)$ and μ , respectively, without much caring about the exact microfoundations.²⁷

The range of the composite coefficients on the two inflation rates could also be extended if, to economize on notation, we leave the economic interpretation of the

²⁶ If $\hat{\theta}^A$, $\hat{\theta}^B$ and the matrices in the above equations remained unchanged, $\sqrt{T'} \text{QLR}/\hat{s} > 1.96$ would obtain if $T' > (1.96/1.51)^2 \cdot T$, i.e. $T' \geq 132$.

²⁷ This is the version that, without discussing further details of its theoretical background, Fuhrer (2006, p. 53) presents as the “canonical hybrid New Keynesian Phillips curve”.

parameter α aside and allow it to exceed unity. This is how we proceed in the present subsection. Formally, the Phillips curve equation in (2.12) need not be altered then. Carrying out the estimation once more without the constraint does indeed drive α further up to a value larger than 2; see model D in Table 2.3. With the estimated $\alpha = 2.214$, the composite coefficient on lagged inflation amounts to 0.69, which is higher than the values that Fuhrer (2006) got from his GMM estimations in a simplified framework but lower than his value of 0.94 from a maximum likelihood estimation of the same coefficient (for a sample eight years longer than our GI period; cf. Fuhrer, 2006, pp.67–69). Although we abstain here from a discussion of the precision of these results, they underline the important role of backward-looking behavior in the firms' price setting even more strongly than before.

Table 2.3: GI estimations (the economic constraints on α and χ are dropped).

Model	α	κ	σ_π	χ	τ	σ_x	ϕ_π	ϕ_x	ϕ_r	J
A	1.000	0.051	0.571	1.000	0.096	0.610	1.482	0.030	0.333	47.6
D	2.214	0.114	0.419	1.000	0.186	0.476	1.115	0.000	0.115	29.6
E	1.000	0.043	0.536	1.460	0.122	0.488	1.616	0.298	0.484	37.7
F	2.484	0.103	0.303	1.574	0.202	0.444	1.606	0.016	0.000	12.7

Note: In all cases, $\rho_\pi = 0$ and $\sigma_r = 0$ results. Values of α and χ exceeding one are admitted for notational convenience; they are not meant to have a meaningful economic interpretation. In model F, the implied coefficients on lagged inflation and lagged output in (2.12) are 0.72 and 0.61, respectively. Bold face figures emphasize the kind of 'excessive' backward-looking behavior admitted in the estimations.

As a somewhat surprising side result we note that the influence of the inherited persistence in the Phillips curve increases, too, rather than decreases, i.e., the estimate of the slope coefficient κ doubles from 0.051 to 0.114. The effect on the entire output-inflation nexus is a simultaneous doubling of τ , the coefficient on the real interest rate in the IS equation.

The improvement in the moment matching to which the higher values of α can give rise is more than only marginal. It is, in particular, remarkable that in the autocovariance diagrams such as those in Figure 2.1, they would now succeed in bringing about a nonnegligible overshooting in all of the nine profiles after their first return to the zero line. Although this reproduces an empirical feature that takes place at lags beyond the horizon of our loss function, the matching over the first eight lags alone diminishes J by already more than one-third, from 47.6 (for $\alpha = 1$) to $J = 29.6$.

Since with respect to the indexation parameter α it proved useful to step outside the original model formulation, we may try the same with the habit parameter χ in

the IS equation, which so far was consistently estimated at its upper bound $\chi = 1$. Reintroducing the upper bound $\alpha = 1$ in the Phillips curve, model E in Table 2.3 shows that also in this way a better fit can be obtained, although with $J = 37.7$ less so than with model D. It is brought about by $\chi = 1.460$, by which the coefficient on lagged output in the IS equation increases from 0.50 to 0.59.

Lastly, it is only natural to drop the constraints simultaneously on both parameters α and χ , which constitutes our model F. The inertia thus made possible do not tend to replace each other but α as well as χ are estimated at similar values to the previous results with only one of the relaxations. Interestingly, no more persistence is now required on the part of the interest rate ($\phi_r = 0$), and the noise levels σ_π and σ_x of the exogenous shocks can subside. Hence the deterministic core of the model gains in importance.

Most remarkable of all, however, is the final improvement in the performance of system (2.12) that is thus achieved. Not only that the two persistence effects from higher values of α and χ do not cancel out, they even reinforce each other. That is, if starting from model A each effect were maintained irrespective of the rest, the value of J would fall to $47.6 - (47.6 - 29.6) - (47.6 - 37.7) = 19.7$. Instead, estimation F reduces the value of the loss function further down to 12.7. With respect to model A this is as strong an improvement as 73%.

While the fit of model A was already fairly good, the fit of model F could therefore be summarized as, we dare say, excellent. The diagrams of the covariance profiles in Figure 2.4 illustrate this to the naked eye. If there still is something to be desired it is a higher variance of the inflation rate in the lower-right panel, and a stronger fall from there to its first-order autocovariance. We would also like to stress that the good matching of the moments considerably extends beyond the 8-lag horizon of the estimation itself.

Despite our excitement about the close fit of estimation F, it is yet another question if F can be said to be significantly better than the other estimations. Here, if anything, F should significantly outperform estimation B with its high value of $J = 119.4$ for the loss function when fixing α at zero. For this comparison, B can be regarded as being nested in F.²⁸ Calculating the 95% quantile of distribution (2.16) as $Q_{0.95} = 48.3$, which falls short of the difference in the loss functions $106.7 = (119.4 - 12.7) = T \cdot \text{QLR}$, we do not only know that B and F have significantly different moment vectors, but we can also conclude that model B is significantly inferior to model F; see Table 2.2. Perhaps somewhat surprisingly, the same table shows that (with the analogous procedure to footnote 28) model F is even significantly better than model A, which previously seemed so satisfactory.²⁹ These

²⁸ Model B has fixed parameters $\alpha = 0$, $\chi = 1$ and $\sigma_r = 0$, while model F only treats $\sigma_r = 0$ as a fixed parameter. Fixing the latter is necessary since otherwise the matrices F^I ($I = B, F$) entering the determination of \widehat{V}^I in (2.16) would not be invertible (owing to $\partial m^I / \partial \sigma_r = 0$ at $\sigma_r = 0$; cf. Appendix A4). We should add that even though the restriction $\rho_\pi \geq 0$ is now dropped for model F, the coefficient continues to be estimated at zero. Hence all parameters that are free in B are also free in F.

²⁹ In order to compare F to the modified Bayesian estimation BR' from above (with its slightly

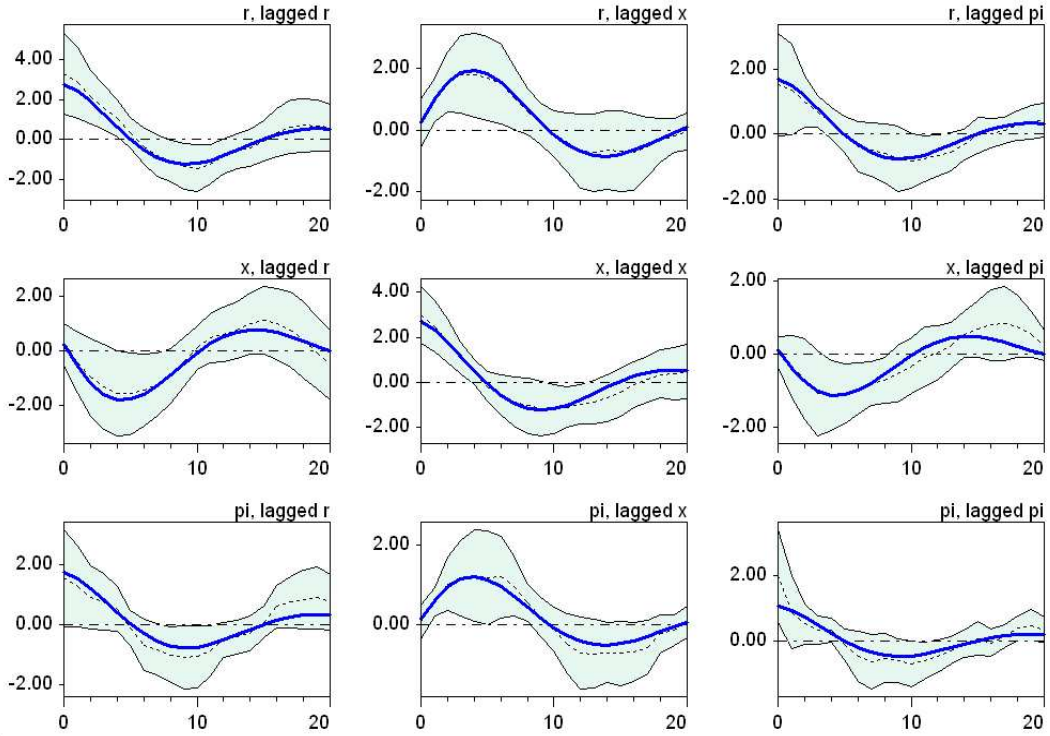


Figure 2.4: Covariance profiles of model F from Table 2.3 (bold lines).

two results give strong emphasis on the beneficial role of backward-looking behavior in the Phillips curve and IS equation, if we adopt a moment matching perspective. Our investigation thus calls for a reconsideration of the microfoundations that would permit the resulting coefficients on lagged inflation and lagged output to become larger than one-half.

2.4.5 Evaluation of the estimated parameters

After temporarily transgressing the interpretational framework for the indexation and habit persistence parameters, we return to our main estimation A with the corner solution $\alpha = 1$ and $\chi = 1$. Let us now have a closer inspection of its parameter estimates. Apart from the issue of the degree of ‘backwardness’ in the Phillips curve and the IS equation, another remarkable result concerns the coefficients in the Taylor rule. Straightforward conventional wisdom has it that over the Great Inflation period the central bank paid (perhaps unduly) strong attention to the variations of economic activity at the cost of price stability, an idea that would be captured by a high policy coefficient ϕ_x on the output gap and a low coefficient ϕ_π on the inflation gap not much above one. This is what we indeed find in the Bayesian

improved fit), where again we want to take advantage of the nested case treatment, we can fix σ_r at $\sigma_r = 0.729$ from BR’ and re-estimate all of the remaining parameters. This gives us estimation F’ (with a deteriorated fit). The test statistics reported in Table 2.2 show that then F’ is significantly superior to BR’.

reference estimation BR in Table 2.1. The moment matching approach, however, reverses the role of the two parameters: the inflation gap coefficient of model A is almost equal to Taylor's famous rule-of-thumb value of 1.50 (Taylor, 1993, p.202), while the output gap takes practically no effect at all. Again in contrast to the Bayesian estimation, with $\phi_r = 0.33$ there is furthermore only weak own-persistence in the rule, which appears all the more surprising as the interest rate inherits no persistence from the output gap.

The complete absence of noise in the monetary policy rule, $\sigma_r = 0$ (something that would imply a stochastic singularity in likelihood estimations), may not be overrated. If for conceptual reasons in a broader context a certain randomness in the conduct of monetary policy were required, we have a wider range over which *ceteris paribus* increases of this parameter have no more than a minimal impact on the loss function, such that in the autocovariance diagrams in Figure 2.1 the human eye would hardly notice any difference. For example, the model-generated variance that the interest rate gap in indirect ways inherits from the other two random shocks is as high as 3.21 for $\sigma_r = 0$, and a rise of σ_r to 0.50 would increase it to just 3.46.³⁰ Technically speaking, σ_r is thus only weakly identified, or white-noise effects in the policy rule have an almost negligible bearing on the overall fit of the model.

The observation on σ_r brings us to the general question of the accuracy of the estimated parameters. As indicated at the end of Section 2.2, we use re-estimations on the model-generated moments to construct 95% confidence intervals for them. Here Hall's method (specified in Appendix A2) serves to obtain the confidence intervals if the parameters are estimated at an interior value (these are the coefficients κ , σ_π , τ , σ_x , ϕ_π , ϕ_r), while the standard percentile intervals are preferred if they are estimated at, or close to, one of the end-points of their admissible range (these are α , ρ_π , χ , ϕ_x and σ_r). A sample size of $B = 1000$ is sufficient for the bootstrap. In this way we arrive at the intervals given in column A of Table 2.1.

Most of the confidence intervals of the MM estimation are wider than those from the Bayesian approach. Apart from σ_r , all of the other parameters are nevertheless reasonably well identified. The frequency distributions of the re-estimated parameters are plotted in the last 11 panels of Figure 2.5,³¹ where the shaded areas indicate a 95% probability mass with the end-points being determined by the standard percentile intervals. In particular, the re-estimations confirm that the polar results $\alpha = 1$, $\rho_\pi = 0$ and $\chi = 1$ are no outliers. Note also that even several intervals in the interior of the admissible range are not symmetric around the estimated parameter values, so that the standard intervals shown here differ from the Hall percentile intervals in Table 2.1. Examples for this are the parameters κ , τ and ϕ_r .

³⁰ A further increase of the noise level up to $\sigma_r = 1.00$, say, would have a stronger effect as it raises the variance to 4.21. Regarding the "the indirect ways" in which the other shocks act on the interest rate, it may be noted that in spite of $\phi_x \approx 0$, a fall of σ_x to zero in the IS equation would cause a drop of $\text{Var}(\hat{r}_t)$ from 3.21 to 1.98. The main reason for this is the fall of $\text{Var}(\hat{\pi}_t)$ from 1.80 to 1.23.

³¹ The density functions are estimated by means of the Epanechnikov kernel; see Davidson and MacKinnon (2004, pp.678–683) for the computational details.

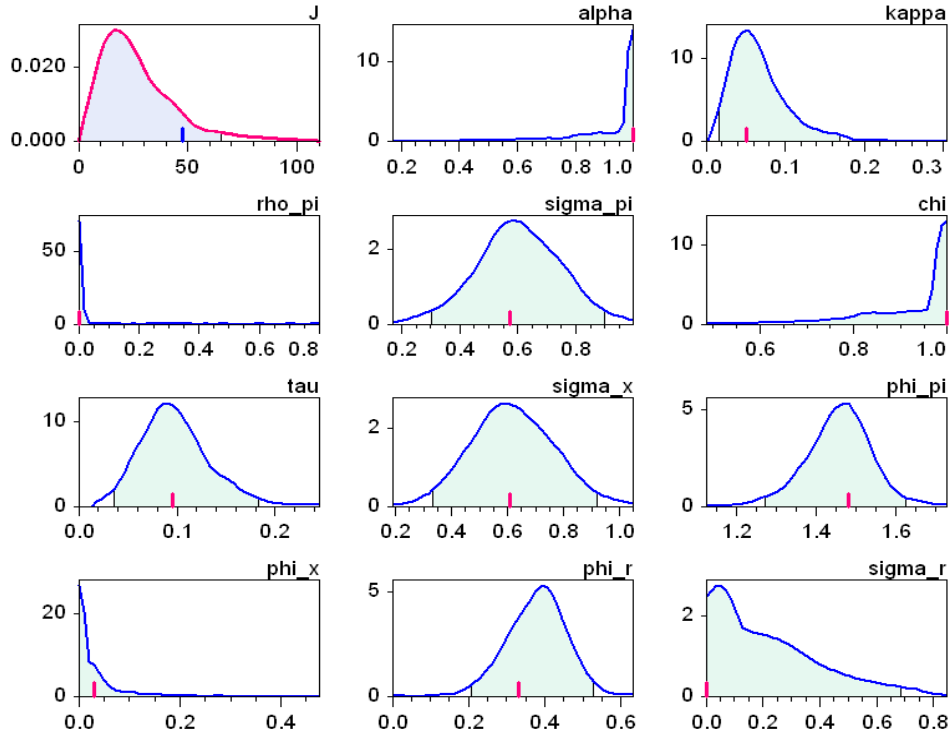


Figure 2.5: Frequency distributions of the re-estimations of the bootstrapped model A (GI).

Note: The bold bars at the bottom indicate the estimates on the empirical moments, the shaded areas show a 95% probability mass of the distributions.

Of course, the re-estimated parameter values are not all independent of each other. On the basis of the discussion of the different sources of persistence in the Phillips curve it will, in particular, be expected that the estimates of α and ρ_π are inversely related. With a negative correlation coefficient of -0.71 , Table 2.4 documents that this is indeed the tightest relationships between two parameters that we can find. As indicated by the bold type numbers there are, however, also other parameters that are closely connected, where most of the pairwise dependencies are within each of the three equations of the model. Two remarkable exceptions are a certain tendency that an increase of κ in the Phillips curve goes along with an increase of τ in the IS equation, and an increase in the supply shock level σ_π (but not persistence ρ_π) in the Phillips curve goes along with an increased persistence ϕ_r in the Taylor rule (and therefore with a decrease in σ_π). The other interdependencies do not seem too surprising and may stand for themselves.

Let us finally turn to the top-left panel of Figure 2.5, which displays the distribution of the minimized values J^b of the loss function in the re-estimations; see eq. (2.11). As indicated by the shaded area, its 95% quantile is $J_{0.95} = 65.0$. The estimated value $\hat{J} = 47.6$ is clearly below this benchmark, so the bootstrap test under

Table 2.4: Pairwise correlations of parameter re-estimates of model A (GI).

	α	κ	ρ_π	σ_π	χ	τ	σ_x	ϕ_π	ϕ_x	ϕ_r	σ_r
α :	1.00	-0.08	-0.71	0.29	0.13	-0.06	0.02	0.05	0.04	0.30	-0.11
κ :		1.00	0.35	0.26	-0.14	0.39	0.07	0.18	0.03	0.18	-0.33
ρ_π :			1.00	-0.35	-0.10	0.11	-0.02	0.04	0.05	-0.08	-0.02
σ_π :				1.00	0.24	0.27	0.03	-0.01	-0.02	0.35	-0.41
χ :					1.00	0.12	-0.44	-0.24	-0.12	0.23	-0.11
τ :						1.00	0.21	-0.14	-0.01	0.00	-0.19
σ_x :							1.00	0.20	0.13	0.01	0.01
ϕ_π :								1.00	0.39	0.28	-0.22
ϕ_x :									1.00	0.47	-0.10
ϕ_r :										1.00	-0.49
σ_r :											1.00

Note: Bold face figures emphasize higher correlation coefficients.

the null hypothesis cannot reject the model. Since the quantile of \hat{J} is 88.38%, the model may be said to have a moment-specific p -value of 11.62%. We nonetheless formulate this only as a conventional statement to succinctly evaluate the overall goodness-of-fit; of course, it is not meant to imply that model A could be the “true” model of the economy.

2.5 The Great Moderation period

In this section we consider the period of the Great Moderation, where in other respects we can proceed along the same lines as above. Our main result is the comparison of estimation A with a Bayesian reference estimation BR in Table 2.5. Again, as in the Great Inflation sample and emphasized by the bold face figures, in contrast to BR estimation A needs no persistence from the shock process in the Phillips curve ($\rho_\pi = 0$), and it yields a high degree of price indexation α , although it is here not maximal.

Apart from that, the general wisdom that inflation during the GM period was less exposed to exogenous shocks than during GI is corroborated by the estimation of the noise level σ_π , which is reduced by almost two-thirds (cf. estimation A in Table 2.1). Also the driving variables in the Phillips curve and the IS equation have a somewhat weaker influence than in GI (lower estimates of κ and τ and narrower confidence intervals). On the other hand, the Taylor rule exhibits stronger persistence ϕ_r . In addition, it is more responsive to the output gap (higher value of ϕ_x), while the estimated coefficient on the inflation gap ϕ_π has a similar order of magnitude to GI. These statements have, however, to be qualified since, in striking contrast to the Bayesian reference estimation BR shown in the first column of Table

Table 2.5: Parameter estimates for GM.

	Bayesian	Moment Matching			
	BR	A	B	C	D
α	0.033 (0.000 – 0.071)	0.816 (0.475 – 1.000)	0.000	0.459	0.863
κ	0.163 (0.103 – 0.221)	0.030 (0.000 – 0.046)	0.139	0.049	0.020
ρ_π	0.389 (0.274 – 0.510)	0.000 (0.000 – 0.453)	0.712	0.000	0.000
σ_π	0.517 (0.420 – 0.611)	0.200 (0.140 – 0.373)	0.176	0.455	0.163
χ	0.825 (0.759 – 0.891)	1.000 (0.669 – 1.000)	1.000	1.000	∞
τ	0.017 (0.009 – 0.025)	0.047 (0.000 – 0.085)	0.045	0.040	0.275
σ_τ	0.346 (0.296 – 0.399)	0.532 (0.295 – 0.702)	0.515	0.504	0.555
ϕ_π	1.181 (1.001 – 1.383)	1.626 (0.295 – 3.746)	2.412	2.784	1.418
ϕ_r	1.014 (0.602 – 1.419)	1.031 (0.176 – 2.129)	0.664	0.687	1.296
ϕ_x	0.814 (0.762 – 0.867)	0.776 (0.673 – 0.958)	0.753	0.786	0.760
σ_r	0.449 (0.395 – 0.502)	0.472 (0.296 – 0.942)	0.527	0.393	0.348
J	170.1	54.1	68.4	72.8	39.6
MCI missed	15	3	4	1	2
p -value	—	5.4%	—	—	—

Note: The discount factor is $\equiv 0.99$ throughout. In estimation B, α is fixed at 0.00. The smaller numbers indicate the confidence intervals; from the posterior distribution in a Bayesian reference estimation (BR), while in estimation A they are computed from (A1) for α , ρ_π , χ , ϕ_π , ϕ_x and from (A2) for κ , σ_π , τ , σ_x , ϕ_r , σ_r . The last row gives the number of moments ('M') that miss the confidence intervals ('CI') of the empirical moments.

2.5, both of these parameter estimates have extremely wide confidence intervals.³² In our moment matching estimation approach we have therefore no firm basis to compare the two policy coefficients ϕ_π and ϕ_x between GI and GM. Incidentally, the width of the confidence intervals is not so much different from the intervals that Cho and Moreno (2006, pp. 1467ff, Tables 2, 4, 5) obtain from their maximum likelihood bootstrap re-estimations of a similar three-equations model (their sample period is 1980:4–2000:1).

The distributions of the re-estimates from the bootstrap for these and the other parameters, on the basis of which the confidence intervals are computed, are shown in Figure 2.6. Note that just as for GI, the distribution of χ strongly leans against one, and the distribution of ρ_π against zero. Regarding the indexation parameter α , the distribution has most of its probability mass not very far below unity (the median is 0.846).

Figure 2.6 is accompanied by the pairwise correlations for these estimates in Table 2.5. Comparing it to Table 2.4 for GI, the following four changes are noteworthy. (1) Not only is the correlation coefficient between κ and τ reduced by one half, but there is now also a positive correlation between α and τ , which was previously negligible. (2) There is a moderate positive correlation between σ_π and ϕ_π , and a moderate negative correlation between σ_π and ϕ_x , both of which were not present in GI. (3) The previously weakly positive connection between the re-estimates of χ and τ has strengthened, and the previously strongly negative connection between χ and σ_x has weakened. (4) While in GI the policy coefficients ϕ_π and ϕ_x were positively correlated, this has become a negative relationship in GM.

Turning to the quality of the match of estimation A in GM, with a minimized value $J = 54.1$ of the loss function *versus* $J = 47.6$ in Table 2.1 it appears slightly worse than estimation A in GI. This impression is confirmed by the moment-specific p -value as it was discussed at the end of Section 2.4.5. In the top-left panel of Figure 2.6, which presents the distribution of the minimized values J^b of the re-estimations, it can be seen that its 95% quantile $J_{0.95}$ almost coincides with the originally estimated \hat{J} . The exact numbers are $J_{0.95} = 55.0$ and $\hat{J} = 54.1$, which constitutes a quantile of 94.6%. The model's p -value therefore amounts to 5.40%, compared to 11.63% for GI.³³

³² Several of the low estimates of ϕ_π might imply indeterminacy with one stable root too many in the Blanchard-Kahn condition. This poses no problem for us since the solution matrix Ω in (2.2) was computed by employing the brute force iteration procedure mentioned in Binder and Pesaran (1996, p. 155, fn 26). First, for the present model even a crude initialization like 0.80 times the identity matrix proves good enough to ensure convergence. Second, in the case of multiple solutions the iteration selects one of the solution matrices automatically and, as we have checked by a number of examples, the most appropriate one—which means that Ω changes continuously when *ceteris paribus* variations of ϕ_π lead the system from determinacy to indeterminacy. By the way, the high robustness of the method is in contrast to the sufficient, somewhat special conditions for local convergence given by Bai et al. (2005, pp. 116f).

³³ Cho and Moreno (2006) evaluate their three-equations model by bootstrapping and re-estimating the model and a low-order unconstrained VAR, from which subsequently a likelihood ratio test statistic can be computed. The resulting p -value is zero for their base model but interestingly, with $p = 3.90\%$ (see their Table 6 on p. 1474, panels A and B) this statistic is not too

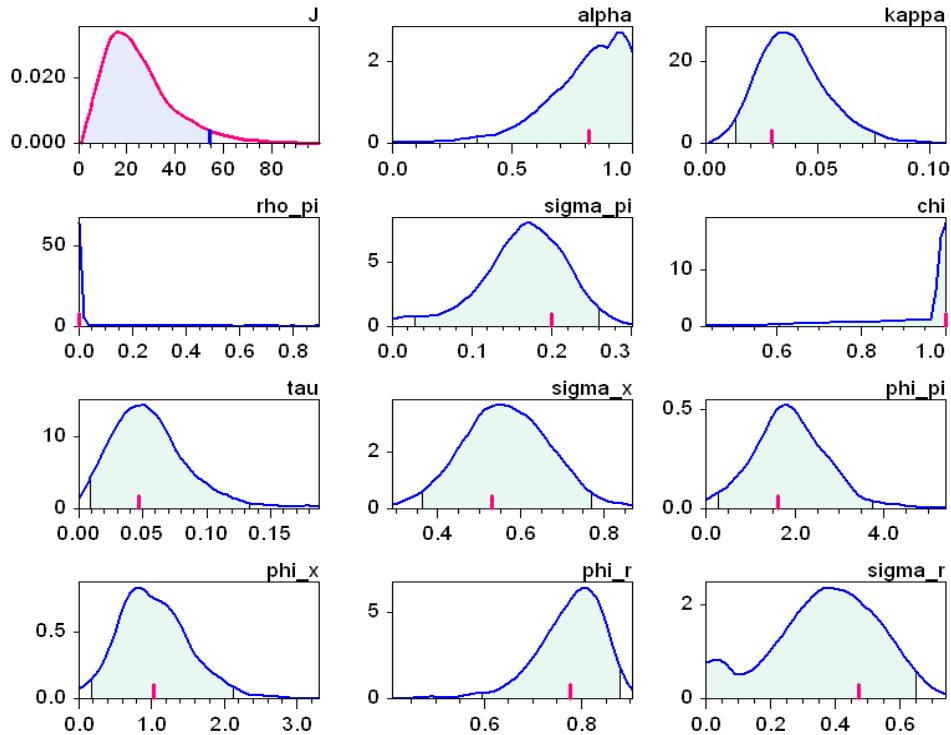


Figure 2.6: Frequency distributions of the re-estimations of the bootstrapped model A (GM).

Note: The bold bars at the bottom indicate the estimates on the empirical moments, the shaded areas show a 95% probability mass of the distributions.

Considering the matching of the single moments, there are now three moments that miss the empirical confidence intervals, *versus* none in GI. Figure 2.7 shows that responsible for this is the steep initial decline of the auto-covariance profile of the inflation gap, which means that in GM there is noticeably less persistence in $\hat{\pi}_t$ than in GI. As it turns out, the model is not too well prepared for that, so that one may be even tempted to say that in its entirety the model tends to exhibit too much, rather than too little, inflation persistence. Specifically, it seeks to find a compromise by first strongly underestimating the level of the variance of $\hat{\pi}_t$, the corresponding t -statistic being -3.17 , and then moderately overestimating $\text{Cov}(\hat{\pi}_t, \hat{\pi}_{t-1})$ and $\text{Cov}(\hat{\pi}_t, \hat{\pi}_{t-2})$ with a t -statistic of 2.30 in both cases; see the bold (red) line in Figure 2.7. In this—but only in this—respect, the Bayesian reference estimation BR (see the dotted (blue) line) proves to be somewhat superior; for the other types of moments, BR displays a similar inferiority to that in GI.

In an attempt to force all of the model-generated moments into the empiri-

different from ours if they admit auto- as well as cross-correlations in all of the random shocks (which on the other hand are features that our estimates can dispense with).

Table 2.6: Pairwise correlations of parameter re-estimates of model A (GM).

	α	κ	ρ_π	σ_π	χ	τ	σ_x	ϕ_π	ϕ_x	ϕ_r	σ_r
α :	1.00	-0.08	-0.62	-0.05	0.22	0.35	0.27	0.11	0.01	0.11	-0.08
κ :		1.00	0.22	-0.12	-0.21	0.20	0.18	-0.03	0.22	0.10	-0.11
ρ_π :			1.00	-0.39	-0.10	-0.17	-0.16	-0.06	0.02	-0.05	0.01
σ_π :				1.00	0.00	0.11	0.03	0.29	-0.22	-0.04	-0.05
χ :					1.00	0.33	-0.16	0.15	-0.04	0.04	-0.08
τ :						1.00	0.45	0.06	-0.03	-0.18	-0.05
σ_x :							1.00	0.09	0.05	0.03	-0.09
ϕ_π :								1.00	-0.26	0.20	-0.11
ϕ_x :									1.00	0.59	-0.22
ϕ_r :										1.00	-0.33
σ_r :											1.00

cal confidence intervals, we also experimented with an *ad-hoc* modification of the present loss function. It is essentially the sum of the skilfully weighted and nonlinearly transformed t -statistics of the single moment deviations $(m_i(\theta) - \widehat{m}_{i,T})$, which tolerate small and medium deviations and heavily penalize t -statistics close to or above 2. However, our effort in thus tuning the function was not fully successful. The best we could achieve is a miss of just one confidence interval, which by the way requires a lower degree of price indexation and still no persistence in the supply shocks. Table 2.5 reports this parameter set as estimation C. It goes without saying that the price for this kind of improvement is a larger deterioration of the original loss function J . The remaining moment that is not satisfactorily matched is again an autocovariance of the inflation gap, this time $\text{Cov}(\widehat{\pi}_t, \widehat{\pi}_{t-4})$ with a t -statistics of -3.63 . This underestimation may nevertheless be considered to be pardonable given the peculiar peaks every four quarters in $\text{Cov}(\widehat{\pi}_t, \widehat{\pi}_{t-h})$, $h = 4, 8, \dots$ (although the data is seasonally adjusted and the phenomenon is completely absent in GI).

After discussing the main estimation A, we can follow the second part of the analysis in Section 2.4.2 (neglecting the more detailed first part for reasons of space). Accordingly, we study the impact of varying degrees of price indexation α on the estimated shock persistence ρ_π and the resulting overall fit of the model. Again including the estimates of κ and ϕ_π in this exercise, Figure 2.8 is obtained. Its main difference from Figure 2.3 for GI is, of course, that the function $\alpha \mapsto \min J$ has an interior minimum, although the performance of the model for $\alpha = 1$ is not much worse. Also to the left of the estimated (i.e. minimizing) α , the deterioration of J is not very dramatic. Actually, the test procedure introduced in Section 2.4.3 tells us that the value $J = 68.4$ for the purely forward-looking case $\alpha = 0$ (which is estimation B in Table 2.5) is not significantly different from $J = 54.1$ for $\widehat{\alpha} = 0.816$ in estimation A. More precisely, as documented in the lower part of Table 2.2, even

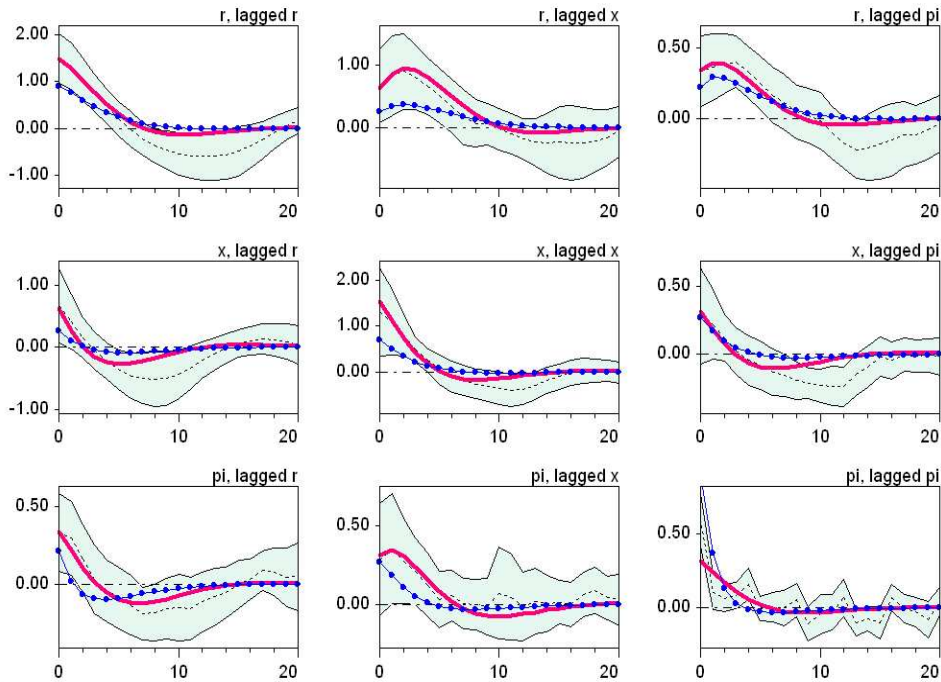


Figure 2.7: estimated *versus* empirical covariance profiles (GM).

Note: The bold (red) line results from the MM estimation A of Table 2.5, the solid (blue) line with dots from the Bayesian reference estimation BR. The shaded area is the 95% confidence band around the empirical moments.

the moments generated by the two estimations cannot be significantly told apart. Incidentally, a comparison of model A with the Bayesian reference estimation leads to the same conclusion.³⁴

Regarding the estimates of ρ_π , κ and ϕ_π that are associated with the exogenous variations in α , Figure 2.8 shares with Figure 2.3 the feature that these parameters are low if α is high and *vice versa*. Again, there is also a discontinuous jump of ρ_π . In Figure 2.8 it is, however, extreme and instead of the monotonic increase of ρ_π as α decreases, there are practically just two states of shock persistence: the estimated ρ_π is zero for $0.64 \leq \alpha \leq 1$, and it marginally falls (rather than increases) from 0.739 to 0.712 as α decreases from 0.63 down to zero. The jump of ρ_π is furthermore so strong that it makes itself also felt in the estimates of κ and ϕ_π .

At the end of this section, we again step outside the interpretational framework of the parameters α and χ and generally admit values exceeding unity for them.

³⁴ Analogously to the treatment for the GI period in Section 2.4.3, BR is modified to BR' by using σ_π as the one and only parameter that is reset to minimize the MM loss function; the new value is then $\sigma_\pi = 0.428$, which reduces the loss from 170.1 to $J = 157.7$.

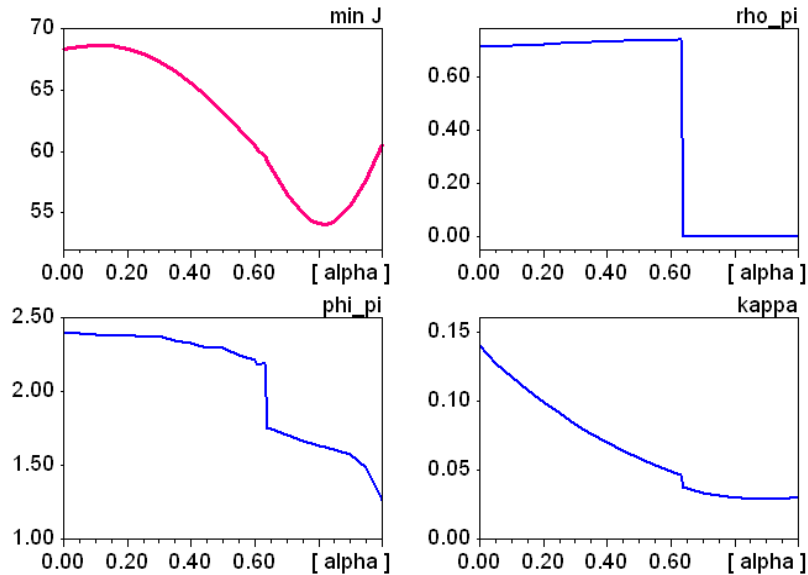


Figure 2.8: MM estimation under exogenous variations of α (GM).

Estimation D in the last column of Table 2.5 shows that the price indexation α makes no use of this option; even if the minimum search procedure for the loss function initializes α considerably above unity, the parameter soon returns into a region of roughly 0.80 or 0.90 (before the other parameters settle down on their final values of the estimation). By contrast, the habit persistence χ strongly tends away from unity, even extremely so. Practically, χ can be said to head towards infinity, which only means that the full weight in the IS equation is on lagged output and the forward-looking component completely disappears. As far as we know, a purely backward-looking IS equation has not yet been obtained in the estimation of New-Keynesian models of similar complexity.

2.6 Conclusion

Being concerned with the estimation of contemporary macroeconomic DSGE models, the main purpose of this chapter was a challenge of the dominant position of the Bayesian approach. Our alternative was the method of moments (MM). In the present application it seeks to match the model-generated second moments of the economic variables to their empirical counterparts, thus summarizing the basic dynamic properties of the model. Besides the relatively low computational cost, a main advantage of the method is its transparency. In this respect, MM allows the researcher to concentrate on what he or she considers to be the most important stylized facts of the economy, and requires him or her to make them explicit. While in the end the choice of moments is a matter of judgement, it is a useful and informative decision to make since a model, at whatever level of complexity, cannot

possibly reproduce all of the empirical regularities that we observe. In addition, the MM approach provides us with an intuitive notion of the goodness-of-fit of a model, which may be checked by visual inspection of suitably organized diagrams or more formally by an econometric assessment of the minimized value of a loss function.

A novel feature of the chapter is that it contrasts the MM with the Bayesian estimation results. To this end we limited ourselves to an elementary three-equations model of the New-Keynesian macroeconomic consensus, where the inflation and interest rates in the structural equations are specified as the deviations from an exogenous flexible trend. Special emphasis was placed on a comparison of the degree of backward-looking behavior in the hybrid Phillips curve. A typical result of many (though not all) Bayesian estimations, to which our framework was no exception, is that lagged inflation tends to play only a minor role in the Phillips curve. Inflation persistence is here brought about by serial correlation in the shock process, besides the inherited persistence from the output gap.

Our MM estimations may add new insights into this discussion. In fact, they found strong evidence to exactly the contrary. With $\alpha \approx 0.80$ the degree of price indexation is high in the Great Moderation (GM) period and it is estimated at its maximal value of $\alpha = 1.00$ in the sample of the Great Inflation (GI), whereas in both cases the supply side shocks are white noise and inherited persistence is weak.

We even took one step further and showed that if, hypothetically, the parameter α were permitted to exceed unity, then in GI it would be as high as almost 2.50. This means that the composite coefficient on lagged inflation in the Phillips curve would be larger than 0.70. The habit persistence parameter χ in the IS equation, by the way, would also be higher than one if it were free in this respect (in both GI and GM).

The much stronger role for the backward-looking elements is all the more important since, already in the presence of the constraints $\alpha \leq 1$ and $\chi \leq 1$, the matching of the empirical moments proves to be fairly good. The general qualitative impression is supported by (moment-specific) p -values above the 5% significance level. Moreover, if the constraints were dropped, the match for GI is so strongly improved that we dared to characterize it as excellent. In that case a new econometric test by Hnatkovska et al. (2009) enabled us to conclude that it is significantly better than our MM benchmark estimation with $\alpha = 1$.

From our perspective there are thus primarily two issues that future research may turn to. First, reconsider the microfoundations for lagged inflation and output in the Phillips curve and IS equation, which still are arguably *ad hoc*—if they at all allow for coefficients on these variables that are larger than one-half.³⁵ Second, apply the MM approach to models with a richer theoretical structure, which would also extend the scope for the moments entering the estimations. The obvious question

³⁵ For the *ad hoc* nature of the common microfoundations of a hybrid Phillips curve, see Rudd and Whelan (2005, pp.20f), which is the longer version of Rudd and Whelan (2007, p.163, fn 7). An interesting new concept to make the Phillips curve more flexible is the hazard function studied by Sheedy (2010), although it comes at the cost of a more complicated structure of lagged and also expected inflation.

would then be whether or not the present results will survive.

Appendix A: Prior densities of the Bayesian reference estimation

The prior densities are essentially taken over from Castelnuovo (2010). One exception is that we mistrust his relatively high estimate of the policy parameter ϕ_π in the GI period, the posterior mode of which, guided by his prior normal distribution around 1.70, amounts to more than 1.80. Following the results by Lubik and Schorfheide (2007) and Benati and Surico (2009), we prefer a lower prior mean and decide on $\phi_\pi \sim (1.3, 0.2)$ for this distribution.

Regarding the prior for the price indexation parameter α we cannot draw on Castelnuovo since, basically (apart from some other specification details), he alternatively fixes α either at zero or one. As his results, like the ones by Ireland (2007) and Cogley and Sbordone (2008) mentioned in the text, favour the purely forward-looking Phillips curve with $\alpha = 0$, we choose a prior mean less than 0.50 but still with some scope for α to move to higher values in the estimation process. So we assume $\alpha \sim \beta(0.3, 0.2)$. Nevertheless, as reported in both Table 2.1 and 2.5, with this setting our estimations show a strong tendency, too, for α to lean against zero. To be self-contained, the priors are all listed in Table A2.6.³⁶ We checked that the posterior densities to which they give rise are in fact well-behaved. This concerns their relationship to the prior densities as well as the convergence checks by Brooks and Gelman (1998), which are summarized in the uni- and multivariate diagnostics provided by Dynare.

Table 2.7: Prior densities of the BR estimations in Tables 2.1 and 2.5.

α	κ	ρ_π	σ_π
$\beta(0.3, 0.2)$	$\Gamma(0.4, 0.1)$	$\beta(0.6, 0.1)$	$I\Gamma(1.0, 8.0)$
χ	τ	—	σ_x
$\beta(0.5, 0.1)$	$\Gamma(0.037, 0.0125)$	—	$I\Gamma(0.25, 2.0)$
ϕ_π	ϕ_x	ϕ_r	σ_r
$\beta(1.3, 0.2)$	$\Gamma(1.2, 0.8)$	$\beta(0.5, 0.28)$	$I\Gamma(1.0, 8.0)$

³⁶ Note that our rates of interest and inflation are annualized, while Castelnuovo's are not.

Appendix B: The standard percentile and Hall's percentile confidence interval

Let a collection $\{\hat{\theta}^b : b = 1, \dots, B\}$ of parameter re-estimates be given, as stated in (2.10). With respect to a significance level $\alpha = 0.05$, let $\hat{\theta}_{i,L}$ be the estimate from (2.10) such that only a fraction $\alpha/2$ of all the bootstrap estimates $\hat{\theta}_i^b$ are less than this value, and likewise $\hat{\theta}_{i,H}$ the estimate that is exceeded by only $\alpha/2$ of the bootstrap estimates. The standard percentile confidence interval is then given by

$$CI_S(\theta_i) = [\hat{\theta}_{i,L}, \hat{\theta}_{i,H}] \quad (\text{A1})$$

(the index S indicating that (A1) is regarded as the standard method.) If the original estimate $\hat{\theta}_i$ from (2.8) lies on the boundary of the admissible set of the parameters, Θ , and $\hat{\theta}_{i,L}$ (or $\hat{\theta}_{i,H}$) coincides with it, then $\hat{\theta}_{i,H}$ (or $\hat{\theta}_{i,L}$) itself will be the $(1 - \alpha/2)$ -quantile (the $\alpha/2$ -quantile, respectively).

Although (A1) is a straightforward specification, it has to be taken into account that it may not have the desired coverage probability. In particular, if $\hat{\theta}_i$ is a biased estimate of θ_i^o , the bootstrap distribution may be asymptotically centred around θ_i^o plus a bias term and, hence, $CI_S(\theta_i)$ is a $(1 - \alpha)\%$ confidence interval for the latter quantity and may thus have a grossly distorted range as a confidence interval for θ_i^o .

An alternative to (A1) that fixes this problem is Hall's percentile confidence interval, which essentially is defined as

$$[2\hat{\theta}_i - \hat{\theta}_{i,H}, 2\hat{\theta}_i - \hat{\theta}_{i,L}] \quad (\text{A2})$$

It is based on the idea that the bootstrap distribution $(\hat{\theta}_i^o - \hat{\theta}_i)$ approximates the distribution $(\hat{\theta}_i - \theta_i^o)$. This implies that $\text{Prob}(\hat{\theta}_{i,L} - \hat{\theta}_i < \hat{\theta}_i - \theta_i^o < \hat{\theta}_{i,H} - \hat{\theta}_i) \approx \text{Prob}(\hat{\theta}_{i,L} - \hat{\theta}_i < \hat{\theta}_i^o - \hat{\theta}_i < \hat{\theta}_{i,H} - \hat{\theta}_i) = 1 - \alpha$, and the first probability expression is easily seen to be equal to $\text{Prob}(2\hat{\theta}_i - \hat{\theta}_{i,H} < \theta_i^o < 2\hat{\theta}_i - \hat{\theta}_{i,L}) = \text{Prob}(\theta_i^o \in CI_H(\theta_i))$. Hence Hall's percentile method is asymptotically correct.

It can, however, happen that $2\hat{\theta}_i - \hat{\theta}_{i,H}$ falls short of a lower bound $\theta_{i,aL}$ of the admissible range of the parameter (something which by construction is not possible with the standard percentile interval). The lower end of the confidence interval may then be set equal to $\theta_{i,aL}$. Similarly so if $2\hat{\theta}_i - \hat{\theta}_{i,L}$ exceeds an upper bound $\theta_{i,aH}$ of the admissible range. We leave such a modification of (A2) aside since in these cases it seems more meaningful to resort to (A1).

Appendix C: Estimation of the moment covariance matrix

$\hat{\Sigma}_m$

Let p_t , q_t stand for the empirical interest rate (gap) r_t , the output gap x_t or the inflation (gap) π_t , as the case may be (the hat on r and π is here omitted). The theoretical covariance of p_t and q_{t-h} is given by $E[(p_t - Ep_t)(q_{t-h} - Eq_t)] = E(p_t q_{t-h}) - (Ep_t)(Eq_{t-h}) = E(p_t q_{t-h}) - (Ep_t)(Eq_t)$. Correspondingly, with respect

to a sample period of length T , we specify the empirical covariance $\text{Cov}(p_t, q_{t-h})$ as being equal to the time average of the products $p_t q_{t-h}$ minus the product of the time averages of p_t and q_t . For the n_m covariances of interest, let there be a total of n_a such average values involved and collect them in a vector $\hat{a} \in R^{n_a}$. For a suitable function $g(\cdot)$ defined on (a subset of) R^{n_a} and attaining values in (a subset of) R^{n_m} , the empirical moments can be expressed as

$$\hat{m}_T = g(\hat{a}) \quad (\text{A3})$$

In order to obtain the covariance matrix of the moments, we first estimate the covariance matrix of the average values \hat{a} . If z_t is a vector the components of which contain all of the lags h of r_t, y_t, π_t that we need ($h = 0, 1, \dots, H$), and $f_j(\cdot)$ for $j = 1, \dots, n_a$ are suitable real functions (to be detailed in a moment) that are defined on these stretches z_t , the time averages can be written as being given by

$$\hat{a}_j = \frac{1}{T} \sum_{t=1}^T f_j(z_t), \quad j = 1, \dots, n_a \quad (\text{A4})$$

While a^o is the ‘true’ value of the real-world data generation process, the vector of its estimates \hat{a} is distributed around it as

$$\sqrt{T}(\hat{a} - a^o) \stackrel{a}{\sim} N(0, \Sigma_a) \quad (\text{A5})$$

For some suitable lag length p (the usual symbol, not to be confused with the above p_t representing r_t, x_t or π_t), a common HAC estimator of the covariance matrix Σ_a is the following ($n_a \times n_a$) Newey-West matrix,

$$\begin{aligned} \hat{\Sigma}_a &= \hat{C}(0) + \sum_{h=1}^p \left(1 - \frac{h}{p+1}\right) [\hat{C}(h) + \hat{C}(h)'] \\ \hat{C}(h) &= \frac{1}{T} \sum_{t=h+1}^T [f(z_t) - \hat{a}] [f(z_{t-h}) - \hat{a}]', \quad h = 0, 1, \dots, p \end{aligned} \quad (\text{A6})$$

Specifically, we follow the advice in Davidson and MacKinnon (2004, p.364) and scale p with $T^{1/3}$. Accordingly we may set $p = 5$ for the two subsamples of the Great Inflation and Great Moderation.

Next, put $m^o = g(a^o)$ and $G_o = [\partial g_i(a^o)/\partial a_j] \in R^{n_m \times n_a}$. Employing the delta method (*cf.* Davidson and MacKinnon, 2004, pp.207f), we know that asymptotically

$$\sqrt{T}(\hat{m}_T - m^o) \stackrel{a}{\sim} N(0, G_o \Sigma_a G_o') \quad (\text{A7})$$

Thus, on the basis of (A6) and the estimated matrix of the partial derivatives \hat{G} , which is constituted by the elements $\partial g_i(\hat{a})/\partial a_j$, the ($n_m \times n_m$) covariance matrix of the moments \hat{m}_T from the finite sample $\{z_t\}_{t=1}^T$ can be estimated as

$$\hat{\Sigma}_m = \hat{G} \hat{\Sigma}_a \hat{G}' \quad (\text{A8})$$

Entering the calculation of the moments $\text{Cov}(p_t, q_{t-h})$ mentioned in the text ($p, q = r, x, \pi$) are the mean values of the products $p_t q_{t-h}$ and, in addition, the three mean values of r_t, y_t and π_t (as already indicated above). This gives us the dimension $n_a = n_m + 3$. Denoting the mean value of a series p_t by a_p and the means of the products $p_t q_{t-h}$ by $a_{pq}(h)$, the n_m covariances can be written as being given by $\text{Cov}(p_t q_{t-h}) = a_{pq}(h) - a_p a_q$.

There are nine different types of covariance profiles. We organize these moments in nine index sets I_1, \dots, I_9 . They do not all contain the same number of indices since for two distinct variables p and q it has to be taken into account that $\text{Cov}(p_t, q_{t-h})$ is included with the lags $h = 0, 1, \dots, H$ in the objective function, but the reverse covariances $\text{Cov}(q_t, p_{t-h})$ only with lags from $h = 1$ onwards. The first and last index in the index sets and the type of covariances assigned to these sets are detailed in the following table. Besides, it once again makes it clear that with $H = 8$, the total number of moments in the objective function is $n_m = 9(H+1) - 3 = 78$.

Table 2.8: Specification of the index sets.

Cov	set	first index	last index	number of indices	functions $g_i(\cdot)$
$r_t r_{t-h}$	I_1	1	$H+1$	$H+1$	$a_{rr}(h) - a_r^2$
$r_t x_{t-h}$	I_2	$(H+1) + 1$	$2(H+1)$	$H+1$	$a_{rx}(h) - a_r a_x$
$r_t \pi_{t-h}$	I_3	$2(H+1) + 1$	$3(H+1)$	$H+1$	$a_{r\pi}(h) - a_r a_\pi$
$x_t r_{t-h}$	I_4	$3(H+1) + 1$	$4(H+1) - 1$	H	$a_{xr}(h) - a_r a_x$
$x_t x_{t-h}$	I_5	$4(H+1)$	$5(H+1) - 1$	$H+1$	$a_{xx}(h) - a_x^2$
$x_t \pi_{t-h}$	I_6	$5(H+1)$	$6(H+1) - 1$	$H+1$	$a_{x\pi}(h) - a_x a_\pi$
$\pi_t r_{t-h}$	I_7	$6(H+1)$	$7(H+1) - 2$	H	$a_{\pi r}(h) - a_r a_\pi$
$\pi_t x_{t-h}$	I_8	$7(H+1) - 1$	$8(H+1) - 3$	H	$a_{\pi x}(h) - a_x a_\pi$
$\pi_t \pi_{t-h}$	I_9	$8(H+1) - 2$	$9(H+1) - 3$	$H+1$	$a_{\pi\pi}(h) - a_\pi^2$

Regarding the n_a functions $f_i(\cdot)$ in (A4), the first n_m of them are defined in accordance with the pairs of variables that are associated with index i in Table A2.6, that is, $f_1(z_t) = r_t r_{t-0}$, $f_2(z_t) = r_t r_{t-1}$, etc., until $f_{n_m}(z_t) = \pi_t \pi_{t-H}$. The remaining three functions capture the average values of the single variables in the obvious order,

$$f_{n_m+1}(z_t) = r_t, \quad f_{n_m+2}(z_t) = x_t, \quad f_{n_m+3}(z_t) = \pi_t$$

All ingredients are thus available to compute $\widehat{\Sigma}_a$ from (A6).

With $a_1 = a_{rr}(0)$, $a_2 = a_{rr}(1)$, etc., the matrix \widehat{G} can be readily set up from the last column in Table A2.6. For $i, j = 1, \dots, n_m$ we simply have $\partial g_i(\widehat{a})/\partial a_j = 1$ if $i = j$, and the partial derivatives are zero otherwise. The last three columns of \widehat{G} , which are the derivatives with respect to $a_{n_m+1} = a_r$, $a_{n_m+2} = a_x$, $a_{n_m+3} = a_\pi$, are given in Table A2.6. It remains to plug this matrix into eq. (A8) to obtain the covariance matrix $\widehat{\Sigma}_m$ of the estimated moments.

Table 2.9: The last three columns of matrix \widehat{G} .

rows corresponding to	column		
	$n_m + 1$	$n_m + 2$	$n_m + 3$
I_1	$-2\widehat{a}_r$	0	0
I_2	$-\widehat{a}_x$	$-\widehat{a}_r$	0
I_3	$-\widehat{a}_\pi$	0	$-\widehat{a}_r$
I_4	$-\widehat{a}_x$	$-\widehat{a}_r$	0
I_5	0	$-2\widehat{a}_x$	0
I_6	0	$-\widehat{a}_\pi$	$-\widehat{a}_x$
I_7	$-\widehat{a}_\pi$	0	$-\widehat{a}_r$
I_8	0	$-\widehat{a}_\pi$	$-\widehat{a}_x$
I_9	0	0	$-2\widehat{a}_\pi$

Appendix D: Specification of the matrices V^X and V^Y in equation (2.15)

First, compute for each model I ($I = X, Y$) the following matrix F^I , which in the specifications further below will be assumed to be non-singular:

$$F^I = \frac{\partial m^I(\theta^I)'}{\partial \theta^I} W \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} - M^I$$

$$M^I = \{E_I \otimes [(\widehat{m}_T - m^I(\theta^I))' W]\} \frac{\partial}{\partial \theta^{I'}} \text{vec} \left[\frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} \right]$$

It is understood that the derivatives are evaluated at the estimated parameter vector $\widehat{\theta}^I$ (we currently omit the hat). These derivatives are well-defined since in the present context only those parameters are treated as free parameters the estimated values of which happen to be in the interior of the admissible set.³⁷ Letting n_θ^I be the dimension of the vector of the free parameters in model I , E_I is here the $n_\theta^I \times n_\theta^I$ identity matrix. The matrices $\partial m^{I'} / \partial \theta^I$ and $\partial m^I / \partial \theta^{I'}$ have format $n_\theta^I \times n_m$ and $n_m \times n_\theta^I$, respectively, so that F^I and M^I are $n_\theta^I \times n_\theta^I$ square matrices. The format of M^I derives from the fact that the matrix in square brackets is a $(1 \times n_m)$ row vector, so that the matrix in curly brackets from the Kronecker product is $n_\theta^I \times (n_\theta^I \cdot n_m)$, while the matrix of the derivative of the vec-expression has the suitable format $(n_\theta^I \cdot n_m) \times n_\theta^I$.

The matrix F^I enters three matrices V_1^I, V_2^I, V_3^I , which are now easily seen to be $n_m \times n_m$ square matrices:

³⁷ For the numerical derivatives the built-in procedures `gradp` and `hessp` in the GAUSS software package are used. The optimal step size for the second derivatives is carefully adjusted because the difference approximations might not be precise when the first derivative is small; see Gill et al. (1981, pp. 127–133) for the details of determining the step size.

$$\begin{aligned}
V_1^I &= \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} (F^{I'})^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} W \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} (F^I)^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} \\
V_2^I &= \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} [(F^{I'})^{-1} + (F^I)^{-1}] \frac{\partial m^I(\theta^I)'}{\partial \theta^I} \\
V_3^I &= \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} (F^{I'})^{-1} (M^{I'} + M^I) (F^I)^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I}
\end{aligned}$$

Finally, the matrices V^I in (2.15) are given by

$$V^I = V_1^I - V_2^I - V_3^I, \quad I = X, Y$$

Structural Estimation of the New-Keynesian Model

This chapter analyzes the empirical relationship between the price-setting/consumer behavior and the sources of persistence in inflation and output. First, a small-scale New-Keynesian model (NKM) is examined using the method of moment and maximum likelihood estimators with US data from 1960 to 2007. Then a formal test compares the fit of two competing specifications in the New-Keynesian Phillips Curve (NKPC) and the IS equation; i.e. backward- and forward-looking behavior. Accordingly, the inclusion of a lagged term in the NKPC and the IS equation improves the fit of the model while offsetting the influence of inherited and extrinsic persistence; it is shown that intrinsic persistence plays a major role in approximating the inflation and output dynamics for the *Great Inflation* period. However, the null hypothesis cannot be rejected at the 5% level for the *Great Moderation* period; i.e. the NKM with purely forward-looking behavior and its hybrid variant are equivalent. Monte Carlo experiments are used to investigate the validity of moment conditions and the finite sample properties of the estimation methods. Finally, the empirical performance of the formal test is discussed along the lines of the Akaike's and the Bayesian information criterion.

3.1 Introduction

In the New-Keynesian model (NKM), some extensions such as habit formation and indexing behavior have gained popularity for the ability to fit the macro data well; see Christiano et al. (2005), Smets and Wouters (2003, 2005, 2007), and Rabanal and Rubio-Ramarex (2005). For example, the forward-looking behavior of price indexation has been challenged by macroeconomists over the last decade, because a hybrid variant of the model with backward-looking behavior provides a good approximation of inflation dynamics; see also Galí and Gertler (1999), Fuhrer (1997), Rudd and Whelan (2005, 2006). In the same way, inertial behavior in the dynamics of the output gap can be better explained by the presence of habit persistence in consumption rule; e.g. see Fuhrer (2000). Accordingly, the lagged dynamics in the NKM influence the transmission of shocks to the economy; the backward-looking behavior in the price-setting and consumption rules affects the degree of endogenous persistence in inflation and output. This also implies that a good approximation of the NKM to the data (e.g. the persistence of aggregate macro variables) can provide a potential explanation for the monetary transmission channel to inflation and

output; see Amato and Laubach (2003, 2004) as well as Woodford (2003, Ch.3).

In a small-scale hybrid NKM, however, current inflation and output depend on its expected future and lagged values, which can give rise to a highly non-linear mapping between structural parameters and the objective function during estimation. Because of this, we cannot easily overcome identification problems in the structural model; in other words, the minimization problem in extreme estimators often does not have a unique solution asymptotically; e.g. see Canova and Sala (2009). The purpose of this chapter is to show to what extent classical estimation methods cope with structural parameter estimates and how these can be used to evaluate the model's empirical performance. Especially, we draw attention to an analytic solution of the model and conduct a structural econometric analysis to identify the effects of a lagged term in inflation and output.¹

More generally, we apply the formal test of Hnatkovska, Marmer and Tang (2012) [HMT henceforth] and examine the significant influence of the lagged term on the inflation and output dynamics. According to HMT, the Vuong-type χ^2 test evaluates the adequacy of a broad class of goodness-of-fit measures and allows for model misspecification; see also Linhart and Zucchini (1986) for model selection. Hence, the test statistic used in our study can simply indicate the goodness-of-fit of the model in hypothesis testing, which measures the discrepancy between the model-generated and empirical moments. For example, Vuong (1989) demonstrates how to use the likelihood ratio test for non-nested models. Rivers and Vuong (2002) generalize the hypothesis testing procedure to empirical applications involving a wide range of estimation techniques. Their procedure extends to complex model selection situations where one or both models may be misspecified and the models may or may not be nested; see Golden (2000, 2003).

The advantage of the formal test of HMT is that the model's empirical performance can be flexibly evaluated according to the chosen moment conditions. The flexibility is commonly associated with the transparency to the fit of the model when the moment conditions are directly binding for parameter estimation. Indeed, the limited information approach has been widely used to estimate parameters of a monetary DSGE model starting from Rotemberg and Woodford (1997). For example, one common approach to this problem is to use impulse responses that are most informative about the DSGE model; Dridi et al. (2007) and Hall et al. (2012) discuss the choice of binding function - i.e. a function which can connect the parameters of the model to the parameters of an auxiliary model - and information criteria for the selection of valid response. Especially, when the model misspecifications and complex structural system do not allow for efficient estimation, the adequacy of the model in fitting the data can be judged by using binding functions; see Gourieroux and Monfort (1995). To conduct empirical analysis without the auxiliary model, Franke et al. (2011) examine a small-scale DSGE model using analytic second moments of the sample auto- and cross-covariances up to lag 8 (two years)

¹Alternatively, the common and simple strategy to provide a quantitative assessment of inflation and output is to use a reduced form (or single equation) estimation, calibration or simulation based inference; see also Gregory and Smith (1991) as well as Nason and Smith (2008).

for estimation as well as model selection. While the empirical results using the moment matching approach are contrasted with the Bayesian estimation, however, the validity of their chosen moment conditions is not indicated by a statistical test.

In this study, we discuss the efficiency of the method of moments (MM) estimation and examine the validity of moment conditions along the lines of the maximum likelihood (ML) approach. To see this, first, we investigate the NKM's empirical performance by using the relationship between interest rate, inflation and output of US data. In particular, we attempt to assess the significance of the lagged dynamics in inflation and output. From the ML and MM estimations, we pinpoint an empirical link between the hybrid model structure and the persistence in inflation and output. Next, the empirical performances of the model with purely forward-looking behavior and its hybrid variant are evaluated using the model selection criterion. Accordingly, the inclusion of a lagged term in the New-Keynesian Phillips Curve (NKPC) and the IS equation improves the fit of the model while offsetting the influence of inherited and extrinsic persistence; it is shown that intrinsic persistence plays a major role in approximating inflation and output dynamics for the *Great Inflation* period. However, the null hypothesis cannot be rejected at the 5% level for the *Great Moderation* period; i.e. the NKM with purely forward-looking behavior and its hybrid variant are equivalent. Finally, we carry out a Monte Carlo study to examine the statistical efficiency of the estimation methods.

This chapter is organized as follows: Section 2 reviews the standard New-Keynesian three-equations model and examines the importance of intrinsic persistence (or backward-looking behavior) for the co-movement between inflation and output. Estimation methodologies and model selection procedures are described in section 3. Section 4 presents the empirical results and the model comparison between the NKM with the forward-looking behavior and its hybrid variant. Moreover, the finite sample properties of the estimators are investigated using the Monte Carlo experiments in section 5. Finally, section 6 concludes. All technical details are collected in the appendix.

3.2 Expectation formation in a DSGE model

In this section, we present the standard New-Keynesian model featuring aggregate supply, aggregate demand (IS), and monetary policy equations.² We explore the model specifications of the lagged dynamics in the NKPC and the IS equation, with a focus on the backward- and forward-looking behavior.

²Smets and Wouters (2003, 2007) empirically examine a medium-scale version of the NKM. They estimate structural parameters and idiosyncratic shocks with the Bayesian techniques. In our study, however, we study a small-scale general equilibrium model and investigate the role of optimizing behavior in the dynamics of inflation and output.

3.2.1 The New-Keynesian three-equations model

Microfoundations of supply- and demand-side economy have been established as the key components of a New-Keynesian model framework; e.g. the behavior of optimizing economic agents. The monetary policy behavior is described by the Taylor rule where the lagged interest rate reflects the gradual adjustment of a central bank. Thus the model is a convenient tool for modeling systemic changes in the economy. Especially, in our current study, we attempt to examine to what extent the gaps of interest rate, inflation and output are related to each other and to what extent they affect the economy ($\hat{\pi}_t := \pi_t - \pi_t^*$, $\hat{r}_t := r_t - r_t^*$). The trend components of the quarterly data are estimated by using the Hodrick-Prescott filter with the smoothing parameter of $\lambda=1600$.³ The standard model reads as follows:

$$\begin{aligned}\hat{\pi}_t &= \frac{\beta}{1 + \alpha\beta} E_t \hat{\pi}_{t+1} + \frac{\alpha}{1 + \alpha\beta} \hat{\pi}_{t-1} + \kappa x_t + \nu_{\pi,t} \\ x_t &= \frac{1}{1 + \chi} E_t x_{t+1} + \frac{\chi}{1 + \chi} x_{t-1} - \tau (\hat{r}_t - E_t \hat{\pi}_{t+1}) + \nu_{x,t} \quad (3.1) \\ \hat{r}_t &= \phi_r \hat{r}_{t-1} + (1 - \phi_r) (\phi_\pi \hat{\pi}_t + \phi_x x_t) + \varepsilon_{r,t} \\ \nu_{\pi,t} &= \rho_\pi \nu_{\pi,t-1} + \varepsilon_{\pi,t} \quad (\text{for indexing behavior}) \quad (3.2) \\ \nu_{x,t} &= \rho_x \nu_{x,t-1} + \varepsilon_{x,t} \quad (\text{for consumption behavior})\end{aligned}$$

where the variable x_t is the output gap, $\hat{\pi}_t$ is the inflation gap and \hat{r}_t is the interest rate gap. The discount factor and the slope coefficient of the Phillips curve are denoted by the parameters β and κ , respectively. The parameters α and χ measure the degree of price indexation in the NKPC ($0 \leq \alpha \leq 1$) and habit persistence of the household ($0 \leq \chi \leq 1$). And τ is a parameter that refers to the intertemporal elasticity of substitution of consumption ($\tau \geq 0$). In the Taylor rule, ϕ_r determines the degree of interest rate smoothing ($0 \leq \phi_r \leq 1$). The other parameters ϕ_x and ϕ_π are the policy coefficients that measure the central bank's reactions to contemporaneous output and inflation ($\phi_x, \phi_\pi \geq 0$).

The shocks $\varepsilon_{z,t}$ are normally distributed with standard deviation σ_z (i.i.d. with $z = \pi, x, r$). Since $\nu_{\pi,t}$ and $\nu_{x,t}$ are autoregressive processes, the persistence of the cost-push and demand shocks are captured by the parameters ρ_π and ρ_x , respectively ($0 \leq \rho_\pi, \rho_x \leq 1$). In estimation, we do not take them together, but treat them as being an independent case in order to directly disentangle the sources of inflation and output persistence in the model.⁴

³Note here that we use the gaps instead of the levels for interest rate and inflation. Indeed, many empirical studies provide evidence for a time-varying trend in inflation and the natural rate of interest; see Castelnuovo (2010), Cogley and Sbordone (2008), and Cogley et al. (2010). Moreover, the second moments are chosen to match the data when we estimate the model parameters. As a result, if we would use the non-stationary data without making assumptions about the data generating process, it would cause substantial bias in parameter estimates of the structural model.

⁴In the current study, we do not consider the presence of serially correlated shocks in the

For the sake of simplicity, we present the above structural equations in canonical form. We denote by y_t and ν_t the vector of three observable variables and the shocks respectively: $y_t = (\widehat{\pi}_t, x_t, \widehat{r}_t)'$, $\nu_t = (\nu_{\pi,t}, \nu_{x,t}, 0)'$.

$$\begin{aligned} AE_t y_{t+1} + B y_t + C y_{t-1} + \nu_t &= 0 \\ \nu_t &= N \nu_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon) \end{aligned} \quad (3.3)$$

where the matrices A , B , C , N are defined in appendix B.

To solve the system, we can express the derivation of the solution as the recursive equation with matrices Ω and Φ . First, we use the method of undetermined coefficients to obtain the unique solution of the system under determinacy (i.e., $\phi_\pi \geq 1$). Second, we apply the brute force iteration method of Binder and Pesaran (1995) to numerically evaluate the matrix Ω ; see appendix B for some intermediate steps.

$$\begin{aligned} y_t &= \Omega y_{t-1} + \Phi \nu_t \\ \nu_t &= N \nu_{t-1} + \varepsilon_t \end{aligned} \quad (3.4)$$

From the matrices Ω and Φ , it follows that the contemporaneous and lagged autocovariance process of the model can be computed recursively using the *Yule-Walker* equations; see chapter 2 of Lütkepohl (2005). On the whole, we adjust the notation by changing the dating of the shocks and rewrite Equation (3.4) as

$$\begin{bmatrix} y_t \\ \nu_{t+1} \end{bmatrix} = \begin{bmatrix} \Omega & \Phi \\ 0 & N \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \nu_t \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} \varepsilon_{t+1} \quad (3.5)$$

Moreover, we can transform Equation (3.5) into the law of motion of $z_t = (y_t', \nu_{t+1}')'$. This can be more compactly written as

$$z_t = A_1 z_{t-1} + u_t, \quad u_t \sim N(0, \Sigma_u), \quad \Sigma_u = D \Sigma_\varepsilon D' \quad (3.6)$$

where the matrix A_1 and the covariance matrix Σ_u are functions of the parameter vector θ . The shocks are mapped into the vector of $u_t = D \cdot \varepsilon_{t+1}$ with $D = (0 \ I)'$. The estimation methodologies will be discussed later.

realizations of interest rate. It is assumed here that the shock persistence parameter of interest rate ρ_r is explained by its lagged term with the smoothing parameter ϕ_r . See also Carrillo et al. (2007).

3.2.2 Sources of persistence: backward- and forward-looking behavior

In the study of the model comparison, we put an emphasis on two polar cases of the behavior of economic agents. For example, when the price indexation parameter α is set to zero, it is assumed in the model that expectations are purely forward-looking. In this case, inflation persistence is exclusively driven by the exogenous shock process and inherited persistence from the output gap (see Table 3.1). But allowing it to be a free parameter, we assume that agents in the market can choose naive expectations. As a result, the NKPC is affected by both expected future and lagged inflation. This allows the model to have a degree of inertia in the dynamics, which can provide structural insights on the comovement between inflation and output.

Table 3.1: Sources of persistence in the NKPC and the IS equation

persistence	inflation	output
intrinsic	indexing behavior (α)	habit formation (χ)
extrinsic	AR (1) of the shock (ρ_π)	AR (1) of the shock (ρ_x)
inherited	slope of Phillips curve (κ)	intertemporal substitution (τ)

In the same vein, Table 3.1 shows that we can distinguish between the backward- and forward-looking behavior in the IS equation. As long as each household chooses consumption optimally (i.e., without habit formation $\chi = 0$), the output dynamics in the economy are only driven by the exogenous shock and the inherited persistence. The latter is implied by rational-expectations equilibrium in the intertemporal allocation of consumption. On the contrary, if habit persistence is present in the consumption rule (i.e., χ is now a free parameter), then the output dynamics is endogenously sustained by the optimizing behavior; the inclusion of habit formation in consumption can explain the dependence of the current expenditure on the past level of expenditure. As a result, the NKPC also depends on the lagged term in the IS equation.

In the current study, we aim to disentangle the sources of inflation and output persistence using classical estimation methods. Especially, we investigate the degree of endogenous dynamics in the model with the lagged term. In other words, it can be seen that the inclusion of the backward-looking behavior in the NKPC and the IS equation offsets the effects of the extrinsic and inherited persistence while strengthening the comovement between inflation and output. Note here that we pinpoint the sources of persistence by separately considering AR (1) of the shocks for the price indexing and consumption behavior.⁵

⁵We also investigated the case where two extrinsic persistence are allowed in the model at the same time; i.e. $\rho_\pi, \rho_x \neq 0$. But in this case we found unreliable parameter values, and the sources of persistence are not clearly identified. For example, in inflation persistence, the results

3.3 Estimation methodologies and model selection

In this section, we explain our estimation methodologies, which must be based on the solution of the model: the method of moment and maximum likelihood estimation. And we present a formal testing procedure such that the empirical performance of the models can be compared.

3.3.1 Method of moment and model comparison: HMT (2012)

From the law of motion in Equation (3.6), it follows that a set of the second moments of z_t can be analytically computed. Thus the contemporaneous and lagged autocovariances of the first-order vector-autoregressive (VAR (1)) are given by:

$$\Gamma(h) := E(z_t z_{t-h}') \in R^{K \times K}, \quad K = 2n, \quad h = 0, 1, 2, \dots \quad (3.7)$$

where n is the dimension of the vector of observable variables y_t . Their computation proceeds in two steps. First, $\Gamma(0)$ is obtained from the equation $\Gamma(0) = A_1 \Gamma(0) A_1' + \Sigma_u$, which yields

$$\text{vec}\Gamma(0) = (I_{K^2} - A_1 \otimes A_1)^{-1} \text{vec}\Sigma_u \quad (3.8)$$

where the symbol ' \otimes ' denotes the Kronecker product. The invertibility of the term $I_{K^2} - A_1 \otimes A_1$ is guaranteed, because A_1 is clearly a stable matrix; i.e. $\phi_\pi \geq 1$. Second, the *Yule-Walker* equations are employed, from which we can recursively obtain the lagged autocovariances as

$$\Gamma(h) = A_1 \Gamma(h-1) \quad (3.9)$$

This formula relates to a vector autoregressive process of the model. From Equation (3.9), we can compute analytic second moments of the model, which will be used to match the empirical counterparts during the MM estimation.

For the purposes of comparison between two models (A and B), we must estimate the model parameters by minimizing a weighted objective function (the chosen goodness-of-fit measures):

of the model estimation indicate high degree of intrinsic persistence in the output dynamics ($\chi = 1.0$). This effect will enter into the Phillips curve, and eventually have influence on the inflation dynamics. If the model allows for extrinsic persistence in the supply shock ($\rho_x \neq 0$), then the effect of intrinsic persistence from the output will be offset by extrinsic persistence, and will affect the intrinsic persistence in inflation. These make it inherently difficult to provide a distinct analysis on the identification of the sources of persistence in the model. Therefore we decide to separately consider the AR (1) shocks in inflation and output.

$$J_I(\theta) \equiv \min_{\theta^I \in \Theta} T (\widehat{m}_T - m^I(\theta^I))' W (\widehat{m}_T - m^I(\theta^I)), \quad I = A, B \quad (3.10)$$

where m^I is a vector of moments, and \widehat{m} is a consistent and asymptotically normal estimator of true moments m_0 .⁶

To examine the macroeconomic effects of the expected future and lagged term on the NKPC and the IS equation, we use auto- and cross-covariances at lag 1 (15 moments) from the interest rate gap (\widehat{r}_t), the output gap (x_t), and the inflation rate gap ($\widehat{\pi}_t$); see also appendix A. With reference to the alternative moment conditions, we present a case for the auto- and cross-covariances up to lag 4 (42 moments). The empirical results of moment estimates and their robustness will be discussed later. Note here that we use a set of the second moments to evaluate the model's empirical performance and apply a formal test to the model of purely forward-looking behavior and its hybrid variant.

In order to construct the objective function, we must estimate the weight matrix W . Here we simply use the Newey-West estimator (Newey and West (1987)):

$$\widehat{\Sigma}_m = \widehat{\Gamma}_T(0) + \sum_{k=1}^5 \left(\widehat{\Gamma}_T(k) + \widehat{\Gamma}_T(k)' \right) \quad (3.11)$$

where $\widehat{\Gamma}_T(j)$ is $\frac{1}{T} \sum_{t=j+1}^T (m_t - \bar{m})(m_t - \bar{m})'$, and k is the number of lags.⁷ In particular, we ignore off-diagonal elements of the weight matrix and compute the inverse of $\widehat{\Sigma}_m$; i.e. $W = 1/\widehat{\Sigma}_{m,ii}$, $i = 1, \dots, n_m$. The reason for this restriction is two-fold: (i) when the sample size is small, the correlation between the elements of the weight matrix and the second moments is likely to be high; e.g. see Altonji and Segal (1996). (ii) If we consider a large set of the moment conditions up to lag of two or three years, the rows in the weight matrix are correlated to some extent. To avoid the dependence between the moments, we only use the diagonal components of the variance-covariance matrix.

Under the standard regularity conditions, the asymptotic distribution of the parameter estimates is given by:

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) \sim N(0, \Lambda) \quad (3.12)$$

⁶The objective function in Equation (3.10) is multiplied by the sample size T , since this gives an asymptotic χ^2 test statistic for testing the null hypothesis that the moment conditions hold; i.e. the model misspecification.

⁷The lag order is chosen following a simple rule of thumb for sample size ($\sim T^{1/4}$). For the GI and GM data, we have 78 and 99 quarterly observations respectively. Therefore k is set to 5.

where we can numerically compute the covariance matrix Λ using the first derivative of the moments at optimum; i.e. $\Lambda = [(DWD')^{-1}]D'W\Sigma_mWD[(DWD')^{-1}]'$.⁸ Note here that D is a gradient vector of moment functions evaluated at the estimated values:

$$\widehat{D} = \left. \frac{\partial m(\theta; X_T)}{\partial \theta} \right|_{\theta=\widehat{\theta}_T} \quad (3.13)$$

Next, we consider hypotheses comparing the goodness-of-fit of the competing models. The null hypothesis H_0 is that two non-nested models fit the data equally:

$$H_0 : \|W^{1/2}(\widehat{m}_T - m^A(\theta^A))\| - \|W^{1/2}(\widehat{m}_T - m^B(\theta^B))\| = 0 \quad (3.14)$$

The first alternative hypothesis is that model A performs better than model B when

$$H_1 : \|W^{1/2}(\widehat{m}_T - m^A(\theta^A))\| - \|W^{1/2}(\widehat{m}_T - m^B(\theta^B))\| < 0 \quad (3.15)$$

The second alternative hypothesis is that model B performs better than model A when

$$H_2 : \|W^{1/2}(\widehat{m}_T - m^A(\theta^A))\| - \|W^{1/2}(\widehat{m}_T - m^B(\theta^B))\| > 0 \quad (3.16)$$

To carry out the model comparison, we define the quasi-likelihood-ratio (QLR) statistic as

$$\widehat{\text{QLR}} = J^B(\widehat{\theta}^B) - J^A(\widehat{\theta}^A) \quad (3.17)$$

According to HMT, the relationship between two models (A and B) can be defined in terms of the following terminologies: (i) nested, (ii) strictly non-nested and (iii) overlapping models. Especially, if the two models share moment conditions for the data generating process and neither model is nested within the other, we assume that two models are overlapping; note here that the two models can not

⁸If the weight matrix is chosen optimally ($\widehat{W} = \Sigma_m^{-1}$), the estimated covariance matrix Λ becomes $(DWD')^{-1}$; see chapter 1 of Anatolyev and Gospodinov (2011) among others. However, in our study, the estimated confidence bands become wider, because the weighting scheme in the objective function is not optimal.

be identical, because they can still generate different moment conditions by using different numerical values for the parameters. Then we can conduct two sequential procedures of the hypothesis testing á la Vuong (1989). To begin, we compute critical values of the QLR distribution for the first step of the model comparison.⁹ The simulated QLR distribution is defined as the following χ^2 -type formula:

$$Z' \widehat{\Sigma}_m^{1/2} W (V^B - V^A) W \widehat{\Sigma}_m^{1/2} Z, \quad Z \sim N(0, E_{n_m}) \quad (3.18)$$

where Σ is a positive definite covariance matrix of the moment estimates, and Z is drawn from the multivariate (n_m) normal distribution; appendix E defines the n_θ^I by n_θ^I matrix V^I with $I = A, B$ in Equation (3.18). If $\widehat{\text{QLR}}$ exceeds the critical value from a 95% confidence interval, then the null hypothesis is rejected. Next, the second step investigates whether or not the source of the rejection asymptotically comes from the same goodness-of-fit. The suggested test statistic has a standard normal distribution (z):

$$w_0 = 2 \cdot \sqrt{(m^B(\theta^B) - m^A(\theta^A))' W (m^B(\theta^B) - m^A(\theta^A))} \quad (3.19)$$

The standard deviation w_0 measures the uncertainty of the $\widehat{\text{QLR}}$ estimates of two models. Accordingly, the null of the equal fits can be rejected when $\sqrt{T} \cdot \text{QLR}(\widehat{\theta}^B, \widehat{\theta}^A) / \widehat{w}_0 > z_{1-0.05/2}$ in which case A is the preferred model, or $\sqrt{T} \cdot \text{QLR}(\widehat{\theta}^B, \widehat{\theta}^A) / \widehat{w}_0 < -z_{1-0.05/2}$ in which case B is preferred.

3.3.2 Maximum likelihood and model selection

Likelihood inference has been widely used to estimate the parameters of DSGE models over the last decade; see Ireland (2004), Lindé (2005) and others. We briefly summarize the econometric steps for the ML estimation and model selection. From Equation (3.4), we may write that:

$$\begin{aligned} y_t &= \Omega y_{t-1} + \Phi \cdot (N \cdot \nu_{t-1} + \varepsilon_t) \\ &= (\Omega + \Phi N \Phi^{-1}) y_{t-1} - \Phi N \Phi^{-1} \Omega y_{t-2} + \Phi \cdot \varepsilon_t \end{aligned} \quad (3.20)$$

where we define the variable $\Phi \cdot \varepsilon_t$ as η_t . Now we assume that η_t follows a multivariate normal distribution.

⁹Appendix E presents intermediate steps for simulating the QLR distribution. The theoretical QLR distribution is derived from the mean value expansion to a binding function (or moment conditions).

$$\eta_t \sim N(0, \Sigma_\eta), \quad \Sigma_\eta \equiv \Phi \cdot \Sigma_\varepsilon \cdot \Phi' \quad (3.21)$$

Hence we can obtain the following conditional probability for the vector of observable variables y_t :

$$y_t | y_{t-1}, y_{t-2} \sim N((\Omega + \Phi N \Phi^{-1})y_{t-1} - \Phi N \Phi^{-1} \Omega y_{t-2}, \Sigma_\eta) \quad (3.22)$$

Given the normality assumption of shocks and data set, the likelihood function can be constructed as:

$$L(\theta) = -\frac{n \cdot T}{2} \ln(2\pi) - \frac{T}{2} \ln |\Sigma_\eta| - \frac{1}{2} \sum_{t=2}^T \eta_t' \cdot \Sigma_\eta^{-1} \cdot \eta_t \quad (3.23)$$

where n is the dimension of y_t . Finally, we arrive at the ML estimates for the parameter θ by maximizing Equation (3.23):

$$\theta_{ml} = \arg \max_{\theta \in \Theta} L(\theta) \quad (3.24)$$

Under standard regularity conditions, the ML estimation is consistent and asymptotically normal:

$$\sqrt{T}(\hat{\theta}_{ml} - \theta_0) \overset{a}{\sim} N(0, (\Upsilon/T)^{-1}) \quad (3.25)$$

where $\Upsilon = E(\partial^2 L(\theta) / \partial \theta \partial \theta')$ is the information matrix. In our study, Υ is numerically computed using the Hessian matrix of the log likelihood function at optimum. For the purposes of the formal test, we use the well-known approach to model selection, the Akaike information criterion (AIC):

$$\text{AIC} = -\frac{2}{T} \cdot \ln L(\theta) + \frac{2p}{T} \quad (3.26)$$

where p is the dimension of the parameter θ . Then, we choose the model for which AIC is the smallest. As an alternative to the AIC, we also consider the Bayesian information criterion (BIC):

$$\text{BIC} = -\frac{2}{T} \cdot \ln L(\theta) + \frac{p \cdot \ln T}{T} \quad (3.27)$$

where the second term, $p \cdot \ln T$ penalizes the model with additional parameters.

3.4 Empirical application

In this section, we present empirical results of the model using the US data. First, we attempt to disentangle the sources of persistence in inflation and output; we examine the empirical performance of the model using the formal test of HMT. Second, the similarities and dissimilarities between the MM and ML estimations are discussed. Finally, we investigate the validity of extra moment conditions based on the model's empirical performance.

3.4.1 Data

The data we use in this study comprise the GDP price deflator, the real GDP and the federal funds rate. The series are taken from the US model data sets by Ray C. Fair; <http://fairmodel.econ.yale.edu/main3.htm>. The trend rates underlying the gap formulation are treated as exogenously given. The trend from a Hodrick-Prescott (HP) filter is used with the smoothing parameter $\lambda = 1600$. The data set covers the period 1960-2007. Due to the structural break beginning with the appointment of Paul Volcker as chairman of the U.S. Federal Reserve Board, we split data into two sub-samples: the *Great Inflation* (GI, 1960:Q1-1979:Q2) and the *Great Moderation* (GM, 1982:Q4-2007:Q2). The data split in the US economy is standard in most existing empirical works.

3.4.2 Basic results on method of moments estimation and model comparison

In this section, we apply the MM estimation to the model and discuss the importance of the lagged dynamics for inflation and output persistence. Auto- and cross-covariances at lag 1 are used as chosen moment conditions. Next, the model comparison method is used to provide a formal assessment of the performance of competing specifications.

3.4.2.1 Assessing the fit of the model to inflation persistence: 15 moments

A set of second moment conditions is used to assess the performance of the two models to fit inflation persistence in the GI data. Table 3.2 shows the parameter estimates for the model with forward-looking behavior and its hybrid variant. As long as the profit maximizing rule without indexation to past inflation (or purely forward-looking) determines the total amount of output in the economy, the inflation dynamics are primarily captured by inherited and extrinsic persistence. Indeed, from the model with purely forward-looking behavior, we obtain much higher estimated values for the parameters κ and ρ_π than its hybrid variant; i.e. $\hat{\kappa} = 0.12$ (forward) > 0.05 (hybrid), $\hat{\rho}_\pi = 0.51$ (forward) > 0.0 (hybrid).

Table 3.2: Parameter estimates for inflation persistence with 15 moments

	GI		GM	
	hybrid	forward	hybrid	forward
α	0.768 (0.007 - 1.000)	0.0 (fixed) (-)	0.105 (0.000 - 1.000)	0.0 (fixed) (-)
κ	0.047 (0.009 - 0.084)	0.123 (0.000 - 0.318)	0.052 (0.000 - 0.136)	0.058 (0.008 - 0.107)
ρ_π	0.000 (-)	0.506 (0.078 - 0.933)	0.000 (-)	0.086 (0.000 - 0.269)
σ_π	0.679 (0.103 - 1.255)	0.778 (0.603 - 0.952)	0.638 (0.454 - 0.823)	0.644 (0.491 - 0.798)
χ	1.000 (-)	0.999 (0.441 - 1.000)	0.774 (0.497 - 1.000)	0.802 (0.499 - 1.000)
τ	0.094 (0.015 - 0.174)	0.089 (0.000 - 0.192)	0.000 (-)	0.000 (-)
ρ_x	0.0 (fixed) (-)	0.0 (fixed) (-)	0.0 (fixed) (-)	0.0 (fixed) (-)
σ_x	0.727 (0.547 - 0.907)	0.662 (0.416 - 0.909)	0.404 (0.118 - 0.691)	0.369 (0.068 - 0.671)
ϕ_π	1.659 (1.000 - 2.334)	1.744 (1.084 - 2.404)	1.798 (1.000 - 4.039)	1.943 (1.000 - 4.465)
ϕ_x	0.378 (0.026 - 0.731)	0.181 (0.000 - 0.452)	0.729 (0.226 - 1.231)	0.652 (0.087 - 1.217)
ϕ_r	0.544 (0.323 - 0.765)	0.463 (0.248 - 0.678)	0.841 (0.698 - 0.984)	0.849 (0.707 - 0.991)
σ_r	0.786 (0.382 - 1.190)	0.662 (0.155 - 1.169)	0.391 (0.099 - 0.684)	0.384 (0.080 - 0.688)
$J(\theta)$	1.30	3.24	2.26	2.44

Note: The discount factor parameter β is calibrated to 0.99. The 95% asymptotic confidence intervals are given in brackets.

Turning to the formal test, we classify the two models into the nested case. Since the hybrid variant of the model can generate richer dynamics due to the lagged inflation with the price indexation parameter α , it nests the other model; the model with the forward-looking expectations does not allow for the effects of intrinsic persistence on the NKPC.

To test the null hypothesis that the two models have an equal fit to the data, we compare the estimated loss function values ($\hat{J}(\theta)$). We find $QLR = 1.94$. The simulated 1% and 5% critical values for the hypothesis testing are 2.42 and 1.31, respectively; see the left panel of Figure 3 in appendix F. Therefore we reject the null hypothesis at 5% level; the backward-looking behavior plays a significant role in approximating the inflation persistence of the GI.

This finding is summarized in Table 3.3. The results show that the hybrid variant of the model can approximate the inflation dynamics better than the other. The inclusion of the lagged term can almost provide perfect fit to the comovements

between interest rate, inflation and output; e.g. see $\text{Cov}(r_t, x_{t-k})$, $\text{Cov}(x_t, \pi_{t-k})$, $\text{Cov}(\pi_t, r_{t-k})$. However, this result does not indicate that the effects of the inherited and extrinsic persistence alone cannot explain the empirical regularities in the US economy. This point should be clear, since the evaluation of the fit of the nested model is not so bad; the estimated values of auto- and cross-covariances at lag 1 lie within the 95% confidence intervals of the empirical moments. According to the formal test, we can only say that there is a significant difference between model-generated moments of the two model, and the fit of the hybrid variant to the data is superior. Note here that we do not aim to match the auto- and cross-covariances up to higher lags; this will be discussed later.

Table 3.3: Empirical and model-generated moments for inflation persistence: 15 moment conditions

Label	Emp.	95% CI	hybrid	forward	Label	Emp.	95% CI	hybrid	forward
$\text{Var}(\hat{r}_t)$	3.296	1.297-5.296	3.400	3.524	$\text{Cov}(x_t, x_{-1})$	2.523	1.356-3.690	2.365	2.495
$\text{Cov}(\hat{r}_t, \hat{r}_{-1})$	2.886	1.142-4.629	2.572	2.388	$\text{Cov}(x_t, \hat{\pi}_t)$	0.069	-0.415-0.552	0.160	0.236
$\text{Cov}(\hat{r}_t, x_t)$	0.232	-0.611-1.075	0.256	0.270	$\text{Cov}(x_t, \hat{\pi}_{-1})$	-0.350	-1.239-0.539	-0.342	-0.234
$\text{Cov}(\hat{r}_t, x_{-1})$	0.991	0.235-1.746	0.946	0.782	$\text{Cov}(\hat{\pi}_t, \hat{r}_{-1})$	1.288	-0.021-2.597	1.067	0.846
$\text{Cov}(\hat{r}_t, \hat{\pi}_t)$	1.535	-0.026-3.097	1.854	2.155	$\text{Cov}(\hat{\pi}_t, x_{-1})$	0.588	0.199-0.977	0.527	0.442
$\text{Cov}(x_t, \hat{\pi}_{-1})$	1.401	0.038-2.765	1.731	1.714	$\text{Var}(\hat{\pi}_t)$	1.989	0.615-3.364	1.713	1.921
$\text{Cov}(x_t, \hat{r}_{-1})$	-0.450	-1.622-0.722	-0.490	-0.369	$\text{Cov}(\hat{\pi}_t, \hat{\pi}_{-1})$	0.893	-0.216-2.001	1.033	0.789
$\text{Var}(x_t)$	3.001	1.728-4.275	3.191	3.176					

Note: 95% CI means the 95% asymptotic confidence intervals for empirical moments. Bold face figures emphasize a better fit of the hybrid model to the empirical moments.

Next, we consider the same steps for the model comparison using the GM data. However, most parameter estimates of the two models do not differ too much. For example, the estimated value for the price indexation is close to zero in the hybrid variant of the model; i.e. $\hat{\alpha} = 0.105$. Accordingly, the result of the formal test shows that the two models fit the data equally well. We find that the estimated QLR statistic is small: $\text{QLR} = 0.17$. The simulated 1% and 5% criteria for the hypothesis testing are 0.51 and 0.27, respectively; see the right panel of Figure 3 in appendix F. Therefore the null hypothesis cannot be rejected.

To save space, we do not report the model-generated moments for GM. Indeed, when we compare trajectories of the model-generated moments (i.e. hybrid and forward), the model covariance profiles overlap each other. The two models provide a good fit to auto- and cross-covariances at the short lag. In other words, we conclude that the two models are not significantly different at 5% level. We discuss the evaluation of the fit of the model using alternative moment conditions later, because the model has a bad fit to the sample autocovariances up to relatively large lags (two or three years).

3.4.2.2 Assessing the fit of the model to output persistence: 15 moments

Turning to the output dynamics in the IS equation, we estimate the effects of habit persistence on the model. The estimated parameters for the model with or without a habit formation are presented in Table 3.4; in the purely forward-looking behavior χ is set to zero, whereas this parameter is subject to the estimation in the hybrid variant of the model. The MM estimates of the two models have similar values except for the degree of the supply shock σ_x , monetary policy shock σ_r and the Taylor rule coefficient ϕ_π .

Table 3.4: Parameter estimates for output persistence with 15 moments

	GI		GM	
	hybrid	forward	hybrid	forward
α	0.517 (0.044 - 0.990)	0.740 (0.204 - 1.000)	0.039 (0.000 - 0.215)	0.036 (0.000 - 0.205)
κ	0.061 (0.011 - 0.112)	0.066 (0.004 - 0.128)	0.064 (0.000 - 0.130)	0.057 (0.000 - 0.117)
ρ_π	0.0 (fixed) (-)	0.0 (fixed) (-)	0.0 (fixed) (-)	0.0 (fixed) (-)
σ_π	0.876 (0.576 - 1.175)	0.715 (0.447 - 0.983)	0.684 (0.545 - 0.824)	0.687 (0.547 - 0.826)
χ	0.931 (0.000 - 1.000)	0.0 (fixed) (-)	0.585 (0.000 - 1.000)	0.0 (fixed) (-)
τ	0.441 (0.000 - 0.943)	0.422 (0.000 - 0.995)	0.480 (0.000 - 1.223)	0.506 (0.000 - 1.315)
ρ_x	0.914 (0.756 - 1.000)	0.868 (0.725 - 1.000)	0.930 (0.864 - 0.996)	0.941 (0.878 - 1.000)
σ_x	0.214 (0.039 - 0.390)	0.445 (0.154 - 0.736)	0.197 (0.000 - 0.452)	0.218 (0.011 - 0.425)
ϕ_π	1.857 (1.000 - 2.729)	2.256 (1.000 - 3.661)	1.109 (1.000 - 2.395)	1.354 (1.000 - 2.905)
ϕ_x	0.838 (0.227 - 1.449)	0.797 (0.244 - 1.349)	1.526 (0.537 - 2.515)	1.438 (0.464 - 2.412)
ϕ_r	0.725 (0.482 - 0.968)	0.835 (0.681 - 0.989)	0.863 (0.773 - 0.953)	0.898 (0.804 - 0.993)
σ_r	0.695 (0.207 - 1.183)	0.240 (0.000 - 1.326)	0.294 (0.060 - 0.528)	0.215 (0.000 - 0.612)
$J(\theta)$	0.44	1.91	0.40	0.57

Note: The discount factor parameter β is calibrated to 0.99. The 95% asymptotic confidence intervals are given in brackets.

It can be seen from the GI data that the estimated demand shock is two times higher in an optimal consumption behavior without habit persistence than the other model ($\hat{\sigma}_x = 0.45$ (forward) $>$ 0.21 (hybrid)). This result shows that the output dynamics are more or less driven by the high level of the demand shocks when a simple rule of thumb behavior is not allowed in the IS equation. As a result, the persistence from the demand shocks also affects inflation dynamics while offsetting the

effects of inherited persistence. This is indicated by a relatively moderate degree of backward-looking behavior; i.e. $\hat{\alpha} = 0.517$ (hybrid) and 0.740 (forward). Moreover, concerning the hybrid model specification, which allows a fraction of consumers to have a rule of thumb behavior, the estimation results indicate a low value for the monetary coefficients on inflation; i.e. $\hat{\phi}_\pi = 2.26$ (forward) $>$ 1.86 (hybrid). Put differently, central banks react weakly to shocks due to the fact that the transmission of the shocks endogenously affects the output persistence; since the parameter estimates are poorly determined with a large confidence interval, however, we might raise doubts about appropriateness of this implication especially when the sample size is small. The reliability of the parameter estimates will be investigated later via a Monte Carlo study.

Now, we compute the loss function values to apply a formal test to the two specifications in the IS equation. In GI, these values are respectively 0.44 and 1.91 for the model with and without habit formation. The simulated 1% and 5% criteria for the hypothesis testing are 1.89 and 1.08, respectively; see the left panel of Figure 4 in appendix F. Since the estimated value for QLR exceeds the criterion at 5% level, we reject the null hypothesis that the two models are equivalent. This implies that the output dynamics are better approximated by the consumption behavior in a rule of thumb manner. This finding is shown in Table 3.5. For example, the hybrid variant of the model can almost provide perfect fit to the covariance profiles of (r_t, x_{t-k}) , (x_t, x_{t-k}) and (π_t, π_{t-k}) .

Table 3.5: Empirical and model-generated moments for output persistence: 15 moment conditions

Label	Emp.	95% CI	hybrid	forward	Label	Emp.	95% CI	hybrid	forward
$\text{Var}(\hat{r}_t)$	3.296	1.297 ~ 5.296	3.305	3.196	Cov $(\mathbf{x}_t, \mathbf{x}_{-1})$	2.523	1.356 ~ 3.690	2.468	2.187
Cov $(\hat{r}_t, \hat{r}_{-1})$	2.886	1.142 ~ 4.629	2.873	3.041	$\text{Cov}(x_t, \hat{\pi}_t)$	0.069	-0.415 ~ 0.552	0.094	0.073
$\text{Cov}(\hat{r}_t, x_t)$	0.232	-0.611 ~ 1.075	0.164	0.342	$\text{Cov}(x_t, \hat{\pi}_{-1})$	-0.350	-1.239 ~ 0.539	-0.417	-0.368
Cov $(\hat{r}_t, \mathbf{x}_{-1})$	0.991	0.235 ~ 1.746	0.984	0.789	$\text{Cov}(\hat{\pi}_t, \hat{r}_{-1})$	1.288	-0.021 ~ 2.597	1.048	1.025
$\text{Cov}(\hat{r}_t, \hat{\pi}_t)$	1.535	-0.026 ~ 3.097	1.657	1.525	$\text{Cov}(\hat{\pi}_t, \mathbf{x}_{-1})$	0.588	0.199 ~ 0.977	0.578	0.579
$\text{Cov}(x_t, \hat{\pi}_{-1})$	1.401	0.038 ~ 2.765	1.582	1.638	Var $(\hat{\pi}_t)$	1.989	0.615 ~ 3.364	1.907	1.810
Cov $(\mathbf{x}_t, \hat{r}_{-1})$	-0.450	-1.622 ~ 0.722	-0.252	-0.073	Cov $(\hat{\pi}_t, \hat{\pi}_{-1})$	0.893	-0.216 ~ 2.001	0.934	1.109
Var (\mathbf{x}_t)	3.001	1.728 ~ 4.275	3.067	3.331					

Note: 95% CI means the 95% asymptotic confidence intervals for empirical moments. Bold face figures emphasize a better fit of the hybrid model to the empirical moments.

In the period of GM, the parameter estimates for the two models are found to be similar. The difference in the loss function values is small (i.e., $\text{QLR} = 0.17$). The simulated 1% and 5% criteria for the hypothesis testing are 7.58 and 12.37, respectively; see the right panel of Figure 4 in appendix F. We cannot reject the null hypothesis that the two models are equivalent. To save space, we do not report the model-generated moments for the GM period; the covariance profiles of the two models overlap each other.

3.4.3 Basic results on maximum likelihood estimation

For comparison purposes, we present the empirical results of the ML estimation of the NKM; it is known in empirical literature that likelihood inference has satisfactory asymptotic properties when the model is correctly specified. In addition, we will examine the large sample properties and statistical efficiency of the estimators in terms of our choice of moments via extensive Monte Carlo experiments later.

Table 3.6: ML estimates for inflation and output persistence

	inflation persistence		output persistence	
	GI	GM	GI	GM
α	0.446 (0.241 - 0.652)	0.157 (0.149 - 0.164)	α	0.478 (0.230 - 0.726)
κ	0.000 (-)	0.036 (0.034 - 0.037)	κ	0.018 (0.000 - 0.099)
ρ_π	0.000 (-)	0.000 (-)	ρ_π	0.0 (fixed) (-)
σ_π	0.879 (0.740 - 1.019)	0.654 (0.649 - 0.660)	σ_π	0.869 (0.737 - 1.002)
χ	1.000 (-)	0.998 (0.978 - 1.000)	χ	0.281 (0.245 - 0.316)
τ	0.037 (0.001 - 0.073)	0.016 (0.014 - 0.019)	τ	0.081 (0.038 - 0.125)
ρ_x	0.0 (fixed) (-)	0.0 (fixed) (-)	ρ_x	0.808 (0.735 - 0.880)
σ_x	0.523 (0.442 - 0.604)	0.253 (0.252 - 0.255)	σ_x	0.211 (0.174 - 0.248)
ϕ_π	1.353 (1.000 - 2.760)	1.001 (1.000 - 1.112)	ϕ_π	1.394 (1.000 - 2.661)
ϕ_x	1.180 (0.295 - 2.064)	1.275 (1.225 - 1.324)	ϕ_x	1.352 (0.710 - 1.995)
ϕ_r	0.809 (0.690 - 0.927)	0.830 (0.827 - 0.833)	ϕ_r	0.803 (0.754 - 0.852)
σ_r	0.734 (0.618 - 0.850)	0.477 (0.472 - 0.481)	σ_r	0.741 (0.622 - 0.859)
$L(\theta)$	-308.86	-233.99	$L(\theta)$	-309.53

Note: The discount factor parameter β is calibrated to 0.99. The 95% asymptotic confidence intervals are given in brackets.

Table 3.6 shows that ML and MM give somewhat similar parameter estimates to the hybrid variant of the model for inflation persistence.¹⁰ For example, the

¹⁰We also investigated the likelihood ratio test between the model with purely forward-looking behavior and its hybrid variant. However, we found that its likelihood value does not differ too

parameter estimates for the price indexation α are 0.45 and 0.16 for the GI and GM data, respectively. The likelihood inference also provides evidence of (moderate) intrinsic inflation persistence in the model. In other words, the backward-looking behavior in the pricesetting rule accounts for inflation persistence. Moreover, the ML estimation gives a very small value for the slope of the Phillips curve ($\hat{\kappa} = 0.0$ (GI) and 0.04 (GM)); individual firms are likely to be less responsive to changes in economic activity (i.e., the Phillips curve is flat). Hence, inflation dynamics in GI are primarily driven by intrinsic (moderate) and extrinsic (strong) persistence; i.e. $\hat{\alpha} = 0.446$, $\hat{\sigma}_{\pi} = 0.879$.

As far as the output persistence is concerned, we find a slight difference between the ML and MM estimation. For example, the comparison of the estimation results between the two estimators shows that the former gives a much lower value for the habit formation parameter ($\chi=0.28$ and 0.25 for the GI and GM data). Further interesting observation from Table 3.6 is that the ML estimate for the intertemporal elasticity of substitution is found to be much lower ($\tau=0.08$ and 0.03 for the GI and GM data). This implies that output persistence is not best captured by the substitution effects from the Fisher equation.

Overall, the slight difference in estimates can be attributed to the assumption of normality of the shocks; if the model is correctly specified, the ML estimation method is more accurate for estimating the NKM than the MM estimation. Since we consider that the model is possibly misspecified to capture the reality (or the true data generating process is not known), however, MM is likely to be a relevant choice for evaluating the model's goodness-of-fit to the data; the moment matching can provide a closer fit to the sample autocovariance. The statistical efficiency and consistency of the parameter estimation adopted in this study will be investigated via a Monte Carlo study later.

Another important point is that the high dimension of the parameter space can give rise to multiple local optima in the likelihood function. However, in the current study, we have a strong confidence in the identification of the structural parameters, because the convergence of the optimization procedure is examined using different starting values. We found that our empirical results converge to a global optimum and are verified by using a set of different optimization such as iterative minimization, Nelder-Mead simplex, and random search method. In this respect, the structural estimation based on the analytic solution of the system is able to cope with the parameter identification problems in a small-scale hybrid NKM. To make a more systemic investigation on our choice of moments, the next section is devoted to examine the parameter estimation of the model using a large set of moment conditions.

much, because the model with purely forward-looking behavior does not make a drastic change in the parameter estimates. In this case, the formal test based on the likelihood function is not effective. We do not report these results to save space.

3.4.4 Validity of extra moment conditions

In this section, we examine the sensitivity of the MM estimation to the chosen moment conditions. From this investigation, we will find that alternative moment conditions do not induce qualitative changes in the parameter estimation. To make our choice of moment conditions more reliable, we make a case for the vector autoregressive (VAR) model with lag 4 as a reference model; see appendix C for optimal lag selection criteria. Accordingly, we analyze the persistence of the macro data in the U.S. economy using auto- and cross-covariances up to lag 4.

3.4.4.1 Assessing the fit of the model to inflation persistence: 42 moments

With a focus on alternative moment conditions of auto- and cross-covariances up to lag 4 (42 moments), we now estimate two specifications of the NKM: forward-looking ($\alpha = 0$) and hybrid case (i.e. α is a free parameter). In Table 3.7, we find evidence of strong backward-looking behavior in the NKPC; $\hat{\alpha} = 1.0$. Moreover, the MM estimates with a small and large set of moments give qualitatively similar values except for the policy shock parameter ($\sigma_r=0.0$).¹¹ For example, in the model with purely forward-looking behavior, the inherited and extrinsic persistence have a substantial influence on the system, compensating for the absence of intrinsic persistence in the NKPC: $\kappa = 0.155$ (forward) > 0.044 (hybrid), $\rho_\pi = 0.675$ (forward) > 0.0 (hybrid).

Next, we draw attention to the model comparison. In the GI data, we found that the price indexation parameter is a corner solution. Accordingly we treat α as being exogenously fixed at unity, because it is assumed in HMT that the estimated parameters are in the interior of the admissible region (see their assumption 2.5 (b)). Put differently, since we consider the price indexation parameter as being exogenously set to different values, it can be seen that two models are now equally accurate and identical in population. In this respect, we treat two models as being overlapping and apply a two step sequential test for model comparison. On the contrary, the value for the estimated price indexation parameter lies in the interior of the parameter space for fitting the GM data ($\alpha = 0.525$). In this case, the hybrid version of the model nests the one with the purely forward-looking expectations.

In the period of GI, the hybrid variant has a better goodness-of-fit to the data ($J = 11.93$) than the purely forward-looking version of the model ($J = 42.77$). As it is discussed above, the estimated AR (1) coefficient for the cost-push shock has no influence on the hybrid NKPC; $\hat{\rho}_\pi = 0.0$.¹² The results also show that inherited persistence has a smaller impact on the output dynamics in the hybrid variant of

¹¹Indeed, likelihood inference would avoid such an estimate provided that there is a stochastic singularity with zero policy shock (i.e., the likelihood value becomes negative infinity at this point).

¹²The estimated value for the parameter σ_r hit the boundary. This makes the objective function ill-behaved and partial derivatives numerically unstable. We set it to zero and compute the numerical derivatives of the other parameters for the model comparison. See appendix D for the matrix notation.

the model ($\hat{\kappa} = 0.044$). This implies that the persistence is best captured by the backward-looking behavior in the hybrid variant. As a result, we find almost perfect fit to the comovements between inflation and output from the hybrid NKM.

Table 3.7: Parameter estimates for inflation persistence with 42 moments

	GI		GM	
	hybrid	forward	hybrid	forward
α	1.0 (-)	0.0 (fixed) (-)	0.509 (0.126 - 0.924)	0.0 (fixed) (-)
κ	0.044 (0.018 - 0.069)	0.155 (0.000 - 0.395)	0.037 (0.000 - 0.075)	0.102 (0.017 - 0.187)
ρ_π	0.000 (-)	0.675 (0.387 - 0.964)	0.000 (0.000 - 0.813)	0.596 (0.367 - 0.825)
σ_π	0.470 (0.000 - 1.686)	0.518 (0.233 - 0.790)	0.364 (0.048 - 0.680)	0.231 (0.093 - 0.369)
χ	1.0 (-)	1.0 (-)	0.770 (0.515 - 1.000)	0.915 (0.518 - 1.000)
τ	0.092 (0.045 - 0.140)	0.063 (0.008 - 0.118)	0.020 (0.000 - 0.055)	0.027 (0.000 - 0.074)
ρ_x	0.0 (fixed) (-)	0.0 (fixed) (-)	0.0 (fixed) (-)	0.0 (fixed) (-)
σ_x	0.716 (0.462 - 0.970)	0.600 (0.348 - 0.853)	0.547 (0.202 - 0.820)	0.468 (0.185 - 0.751)
ϕ_π	1.740 (1.255 - 2.225)	1.809 (1.221 - 2.397)	2.025 (1.000 - 2.870)	2.218 (1.141 - 3.114)
ϕ_x	0.080 (0.000 - 0.542)	0.157 (0.000 - 0.528)	0.563 (0.216 - 1.059)	0.564 (0.154 - 0.974)
ϕ_r	0.267 (0.000 - 0.905)	0.458 (0.224 - 0.692)	0.765 (0.619 - 0.881)	0.732 (0.592 - 0.872)
σ_r	0.000 (-)	0.000 (-)	0.486 (0.303 - 0.727)	0.545 (0.351 - 0.739)
$J(\theta)$	11.93	42.77	23.97	27.47

Note: The discount factor parameter β is calibrated to 0.99. The 95% asymptotic confidence intervals are given in brackets.

In order to examine the significant difference of the fit of the two models, we subtract the objective function value of purely forward-looking NKM from its hybrid variant; i.e. $QLR = 30.83$. According to the simulated test distribution, critical values for the 99% and 95% confidence intervals are 16.99 and 9.96, respectively (see the left panel of Figure 5 in appendix F). Since the test statistic exceeds the critical value at 5% level, we proceed to take the second step of the hypothesis testing, which asymptotically evaluates the estimated moments of two models from the empirical data.

In the second step of the formal test, we examine the uncertainty of the estimated difference between the two models for evaluating their fit to the data. We compute the plug-in estimate of \hat{w}_0 (2.54). Under the null hypothesis, the test static follows

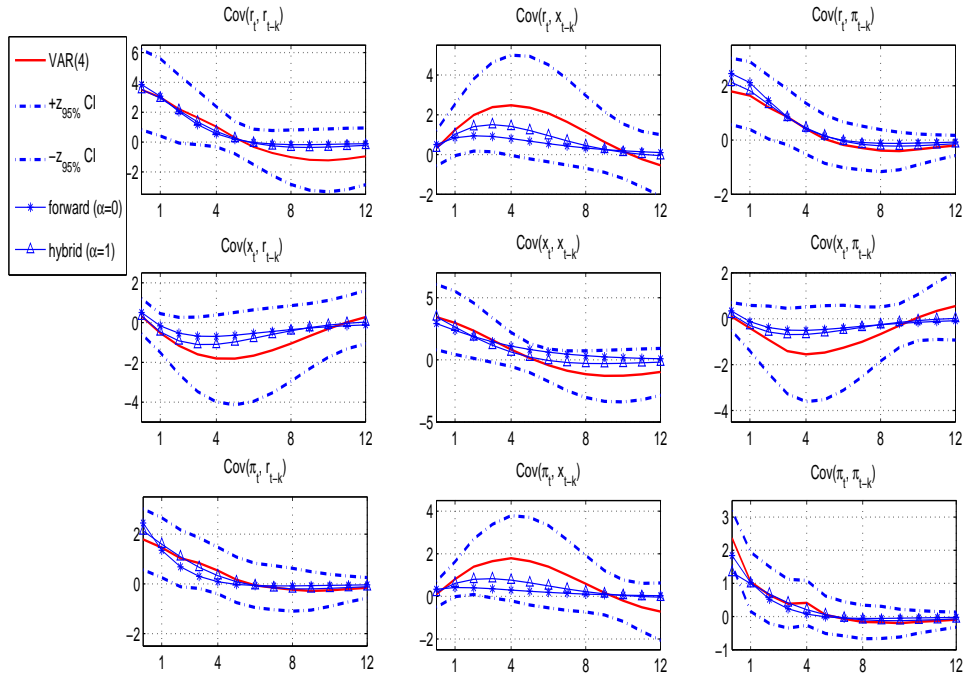


Figure 3.1: Covariance profiles for inflation persistence in GI (dashed: empirical, \triangle : hybrid, $*$: forward)

Note: The empirical auto- and cross-covariances are computed using an unrestricted fourth-order vector autoregression (VAR) model. The asymptotic 95% confidence bands are constructed following Coenen (2005).

a standard normal distribution; i.e. $\sqrt{T} \cdot QLR(\theta^A, \theta^B) \sim N(0, w_0^2)$. The estimate of $\sqrt{T} \cdot QLR/\hat{w}$ is 1.37, which is smaller than a critical value at the 5% significance level of the two-tailed test. Therefore the results show that both models have the same goodness-of-fit to the profile of the empirical moments, and the null hypothesis cannot be rejected.¹³ Figure 1 depicts the model-generated moment conditions at three years for GI and contrasts them with the empirical counterparts of the VAR (4) model. Indeed, a visual inspection of this figure indicates that the two models have different moments, but their matching to the empirical counterparts is not significantly different.

In the period of GM (Table 3.7), it is shown that the hybrid variant fits the data better (23.97). The estimation results provide evidence of the (strong) inherited and

¹³This statistical inference does not remain the same if the price indexation parameter is allowed to exceed unity. The constraint on habit formation parameter (χ) is also removed. See Franke et al. (2011) for details.

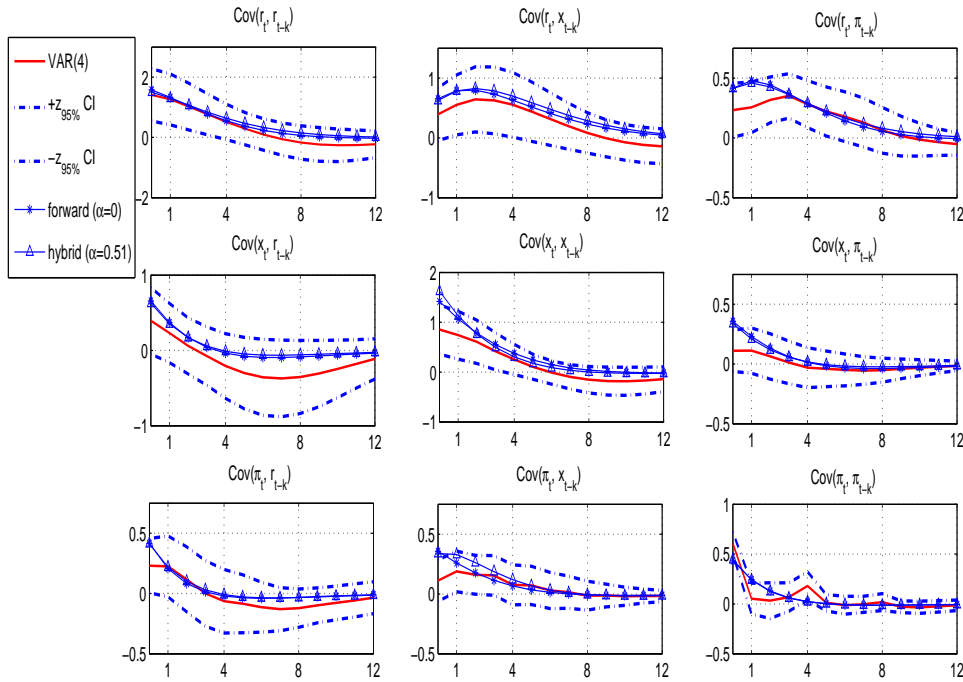


Figure 3.2: Covariance profiles for inflation persistence in GM (dashed: empirical, \triangle : hybrid, $*$: forward)

Note: The empirical auto- and cross-covariances are computed using an unrestricted fourth-order vector autoregression (VAR) model. The asymptotic 95% confidence bands are constructed following Coenen (2005).

extrinsic persistence in the model with purely forward-looking behavior, because these can offset the impact of inherited persistence on the output dynamics; i.e. $\hat{\kappa} = 0.102$ (forward) $>$ 0.037 (hybrid), $\hat{\rho}_\pi = 0.596$ (forward) $>$ 0.0 (hybrid). However, the other parameter estimates are not different in both specifications.

These empirical findings seem to strengthen the relevance of backward-looking behavior for the GM data. However, the difference between the two models (3.49) does not exceed the critical value for the 95% confidence intervals in the formal test; i.e., critical values for 99% and 95% confidence intervals are 38.39 and 21.46, respectively; also see the right panel of Figure 5 in appendix F. Put differently, the effects of intrinsic persistence on inflation can be adequately replaced by the inherited and extrinsic persistence. From this, we cannot identify the sources of the persistence in the system. Therefore we do not proceed to take the second step of the model comparison method and conclude that the null hypothesis cannot be rejected. Figure 3.2 depicts the model-generated moment conditions at three

years for the GM data; the comparison between the model-generated and empirical moments by a VAR (4) process is displayed here.

3.4.4.2 Assessing the fit of the model to output persistence: 42 moments

Table 3.8 reports the MM estimation for the output persistence using alternative moment conditions. The results show that the output dynamics are strongly influenced by the inherited persistence. Indeed, in the case of the intertemporal elasticity of substitution, we obtain high estimated values for the two models: e.g. in GI, $\hat{\tau} = 0.205$ (hybrid), 0.676 (forward). In addition, we find that all the estimated values for ρ_x exceed 0.7. Especially regarding the GI data, this value increases substantially in the model with purely forward-looking expectations, which can cover the absence of intrinsic persistence in the IS equation; i.e. $\chi=0.0$ (fixed), $\hat{\tau} = 0.676$.

Table 3.8: Parameter estimates for output persistence with 42 moments

	GI		GM	
	hybrid	forward	hybrid	forward
α	1.0 (-)	0.998 (-)	0.186 (0.000 - 0.396)	0.203 (0.000 - 0.441)
κ	0.054 (0.005 - 0.102)	0.037 (0.010 - 0.065)	0.086 (0.037 - 0.134)	0.088 (0.027 - 0.149)
ρ_π	0.0 (fixed) (-)	0.0 (fixed) (-)	0.0 (fixed) (-)	0.0 (fixed) (-)
σ_π	0.519 (0.099 - 0.939)	0.428 (0.108 - 0.747)	0.609 (0.461 - 0.757)	0.579 (0.410 - 0.749)
χ	1.0 (-)	0.0 (fixed) (-)	0.991 (-)	0.0 (fixed) (-)
τ	0.205 (0.000 - 0.436)	0.676 (0.000 - 1.897)	0.237 (0.000 - 0.547)	0.236 (0.000 - 0.803)
ρ_x	0.707 (0.290 - 1.000)	0.890 (0.743 - 1.000)	0.854 (0.686 - 1.000)	0.790 (0.583 - 0.997)
σ_x	0.213 (0.016 - 0.410)	0.519 (0.169 - 0.869)	0.140 (0.000 - 0.298)	0.340 (0.037 - 0.642)
ϕ_π	1.741 (1.154 - 2.327)	2.046 (1.000 - 3.134)	2.133 (1.000 - 3.279)	2.224 (1.000 - 3.764)
ϕ_x	0.169 (0.000 - 0.584)	1.103 (0.275 - 1.931)	0.762 (0.189 - 1.335)	0.588 (0.000 - 1.202)
ϕ_r	0.389 (0.000 - 0.853)	0.889 (0.753 - 1.026)	0.770 (0.640 - 0.900)	0.783 (0.648 - 0.917)
σ_r	0.012 (-)	0.016 (-)	0.447 (0.248 - 0.645)	0.448 (0.212 - 0.685)
$J(\theta)$	10.54	31.64	20.79	23.85

Note: The discount factor parameter β is calibrated to 0.99. The 95% asymptotic confidence intervals are given in brackets.

Another point worthwhile mentioning here is that the estimation results of the purely forward-looking model indicate high monetary policy coefficients on interest

rate, inflation and output in GI; i.e. $\hat{\phi}_\pi = 2.05$, $\hat{\phi}_x = 1.10$, $\hat{\phi}_r = 0.89$. Moreover, in the hybrid variant, the parameter χ is almost a corner solution for both the GI and GM data, which strengthens a rule of thumb behavior in consumption. In other words, the rule of thumb behavior reinforces the degree of endogenous persistence in the output dynamics. However, as long as the model predicts that the optimal behavior of household is described by consumption without a simple rule of thumb behavior ($\chi = 0$), the result indicates the strong degree of the demand shocks; the estimated value is more than twice as high as the one of the hybrid model; i.e. $\hat{\sigma}_x = 0.519$ (forward) > 0.213 (hybrid) for GI, 0.340 (forward) > 0.140 (hybrid) for GM.

Turning to the model comparison by using the GI data, we treat the two models as being overlapping, because the habit formation parameter is now a corner solution. In the first step of the model comparison, we compare the objective function values (QLR = 21.10). The simulated 5% and 1% criteria for the hypothesis testing are 19.63 and 34.59, respectively (see the left panel of Figure 6 in appendix F). Since the estimated QLR exceeds the 5% criterion for the model comparison, we support the hypothesis that two models have different moments. In the second step, we estimate $\sqrt{T} \cdot QLR/\hat{w}$ which is 1.02. However, this value does not exceed the criterion in the standard normal distribution. As a result, we conclude that there is no significant difference between two models in matching the empirical moments; i.e. the two models have different moments, but an equivalent fit to the empirical moments. To save space, we do not provide the model covariance profiles for the output persistence. Note here that the result of the MM estimation with a large set of moments provides a closer fit to the sample auto- and cross-covariances up to large lags.

Now we draw attention to the model comparison by using the GM data. To begin, we treat the two models as being a nested case, since the estimated value for the habit formation parameter lies at an interior point. The model without habit persistence is nested within the other. Next, we compute the difference between the objective function values of the two models (QLR = 3.06). Then this value is used to evaluate the null hypothesis of the equal fit of the two models. Since the 5% and 1% criteria for the hypothesis testing are 18.52 and 29.05, respectively (see the right panel of Figure 6), the null hypothesis cannot be rejected. Therefore we conclude that two models have an equal fit to the empirical moments.

In sum, the MM estimation using a large set of moment conditions can provide a stronger evidence for the backward-looking behavior in the pricesetting and consumption rules compared to ML and MM with 15 moment conditions. This result is mainly attributed to the fact that the objective function to minimize is now the norm of additional sample second moments. However, the result of the model comparison becomes inconclusive, because the estimated values for the price indexation and habit persistence parameters were corner solutions; we used the two-step sequential hypothesis testing. We found that the null hypothesis cannot be rejected when the sample size is small. An elaborate analysis of model selection will be discussed in the next section.

3.5 Attaining efficiency from moment conditions

In this section, first, we study the finite sample properties of the estimators for the NKM; in addition, we investigate the effect of model misspecification on the parameter estimation. Second, we discuss the empirical performance of the formal test of HMT along the lines of the Akaike's and the Bayesian information criterion.

3.5.1 Monte Carlo study

The Monte Carlo experiment attempts to clearly demonstrate the statistical efficiency of the estimation methods, which are used in the previous section. In this way, we aim to investigate the role of choice of moments and its influence on the parameter estimation. To begin, we consider the model specification of inflation persistence as the true data generating process; we simulate the artificial economy by using the parameters near to the results of the MM estimation with 15 moments (see Table 3.2): e.g. high degree of backward-looking behavior ($\alpha=0.750$), moderate inherited persistence ($\kappa=0.050$) and no extrinsic persistence ($\rho_\pi=0.0$). Next, we generate 1,000 data sets each consisting of 550 observations. The first 50 observations are removed as a transient period. Three sample sizes are considered: 100, 200 and 500. We use the Matlab R2010a for this study. In optimization, we use the unconstrained minimization "fmincon" with the algorithm 'interior-point'; maximum iteration and tolerance level are set to 500 and 10^{-6} , respectively.

In the Monte Carlo experiments, we consider two cases of model specification; i.e. correctly specified and misspecified. In the former, we discuss the finite sample properties of the MM and ML estimation. Turning to the latter, we consider the model with purely forward-looking expectations and examine the degree of bias in the parameter estimates; i.e. (1) to what extent the extrinsic persistence (ρ_π) is inflated due to the misspecification and (2) to what extent the model misspecification affects the estimates for the other structural parameters.

The main findings for the correctly specified case in Table 3.9 can be summarized as follows:

- The estimate of the price indexation parameter α is downward-biased, whereas the AR (1) coefficient of inflation shocks ρ_π is estimated to be positive. Especially, in ML, we found that the sum of the estimated values for the price indexation and extrinsic persistence parameters is close to the true value of α : i.e. $\alpha + \rho_\pi \approx 0.75$.
- ML has slightly poorer finite sample properties than MM. If the sample size is small, the conventional Gaussian asymptotic approximation to the sample distribution is not as much precise as MM.
- The asymptotic efficiency of the ML estimation appears superior to MM, since the mean of standard errors over 1000 estimations shows that the confidence intervals for the MM estimates are noticeably narrow. However, the large

sample size remarkably improves the asymptotic efficiency of MM with 15 and 42 moments; e.g. $T=500$.

- It can be seen from the Monte Carlo results that the overall parameter uncertainty of MM with 42 moments is higher than ML and MM with 15 moments. However, in this case, MM with 42 moments can provide the most precise estimate on the price indexation parameter α . Note here that the accuracy of statistical inference for the behavior of economic agents (i.e. backward- or forward-looking) comes at the cost of allowing for large uncertainty in the estimates of other structural parameters; in other words, incorporating more second moments in the objective function improves the fit of the model to the persistence of inflation dynamics, but reduces efficiency in the other structural parameters.
- The results using MM with 42 moments show that we obtain the large asymptotic error for the policy shock parameter σ_r ; i.e. S.E = 1.407 for $T=100$. This is attributed to the fact that the estimated values sometimes hit the boundary (i.e. $\sigma_r = 0.0$), which makes the numerical derivative of the moments unstable. This problem does not occur when the large sample size is used (e.g. $T=500$).

Turning to the misspecified case, the results show that there is high correlation between the price indexation and AR (1) coefficient of the supply shocks; see appendix G. Indeed, it is shown in Table 3.13 that the AR (1) coefficient is strongly upward-biased for both MM and ML. The parameter estimates offset the effects of intrinsic persistence on the inflation dynamics; e.g. $\rho_\pi = 0.616$ (ML), 0.632 (MM with 15 moments), 0.598 (MM with 42 moments) when the sample size is 100. The large sample size does not correct the bias of this parameter. Fortunately, the other structural parameters are not influenced by the model misspecification; i.e. we obtain parameter estimates near to the true ones by using both MM and ML. They converge at some reasonable rate towards the true parameters as the sample size gets larger (consistency).

Similarly, the degree of the inflation shock σ_π is more or less downward-biased. In addition, the slope coefficient of the Phillips curve is upward-biased in ML, and the results of the MM estimation show very strong bias: $\hat{\kappa} = 0.096$ (ML), 0.176 (MM with 15 moments), 0.205 (MM with 42 moments) when $T=100$. We can see that (strong) extrinsic and (moderate) inherited persistence offset the absence of intrinsic persistence from the model misspecification. When this result is contrasted by the correctly specified case, however, we obtain a relatively higher value for the estimated objective function: e.g. $\hat{J} = 2.36$ (misspecified) $>$ 0.30 (correctly specified) for $T = 100$, MM with 15 moments. In other words, redirecting the intrinsic persistence to the inherited and extrinsic persistence is not satisfactory enough to provide a good approximation to the data. This is in line with our empirical findings in the previous section; the lagged terms in the NKPC and IS equation are empirically important.

Table 3.9: The Monte Carlo results on the MM and ML estimates, (): root mean square error, S.E : mean of standard error

θ^0		ML			MM with 15 moments			MM with 42 moments		
		T = 100	T = 200	T = 500	T = 100	T = 200	T = 500	T = 100	T = 200	T = 500
α	0.750	0.523 (0.375) S.E : 0.162	0.573 (0.322) S.E : 0.170	0.651 (0.228) S.E : 0.175	0.614 (0.256) S.E : 0.319	0.654 (0.196) S.E : 0.222	0.692 (0.121) S.E : 0.138	0.700 (0.245) S.E : 0.281	0.702 (0.205) S.E : 0.190	0.729 (0.118) S.E : 0.113
κ	0.050	0.074 (0.076) S.E : 0.054	0.066 (0.081) S.E : 0.048	0.056 (0.014) S.E : 0.041	0.083 (0.057) S.E : 0.042	0.068 (0.030) S.E : 0.025	0.058 (0.015) S.E : 0.013	0.093 (0.075) S.E : 0.050	0.073 (0.042) S.E : 0.030	0.058 (0.018) S.E : 0.014
ρ_π	0.000	0.218 (0.330) S.E : 0.112	0.172 (0.284) S.E : 0.1000	0.097 (0.198) S.E : 0.076	0.175 (0.255) S.E : 0.327	0.129 (0.194) S.E : 0.238	0.082 (0.124) S.E : 0.152	0.194 (0.299) S.E : 0.313	0.147 (0.241) S.E : 0.230	0.078 (0.144) S.E : 0.150
σ_π	0.675	0.602 (0.330) S.E : 0.044	0.619 (0.125) S.E : 0.047	0.640 (0.073) S.E : 0.048	0.613 (0.113) S.E : 0.143	0.624 (0.085) S.E : 0.106	0.639 (0.056) S.E : 0.068	0.564 (0.1778) S.E : 0.172	0.584 (0.136) S.E : 0.130	0.618 (0.088) S.E : 0.086
χ	1.000	0.935 (0.113) S.E : 0.159	0.949 (0.090) S.E : 0.183	0.967 (0.053) S.E : 0.201	0.932 (0.108) S.E : 0.173	0.948 (0.078) S.E : 0.126	0.962 (0.055) S.E : 0.082	0.941 (0.075) S.E : 0.207	0.956 (0.083) S.E : 0.151	0.966 (0.059) S.E : 0.098
τ	0.090	0.089 (0.031) S.E : 0.045	0.088 (0.023) S.E : 0.047	0.087 (0.014) S.E : 0.048	0.101 (0.039) S.E : 0.040	0.095 (0.026) S.E : 0.028	0.091 (0.016) S.E : 0.017	0.105 (0.044) S.E : 0.041	0.097 (0.030) S.E : 0.029	0.092 (0.018) S.E : 0.018
σ_x	0.700	0.695 (0.059) S.E : 0.050	0.697 (0.043) S.E : 0.052	0.699 (0.025) S.E : 0.053	0.743 (0.102) S.E : 0.086	0.735 (0.073) S.E : 0.062	0.724 (0.048) S.E : 0.039	0.738 (0.123) S.E : 0.121	0.729 (0.086) S.E : 0.089	0.721 (0.054) S.E : 0.057
ϕ_π	1.650	1.666 (0.183) S.E : 0.345	1.654 (0.118) S.E : 0.316	1.652 (0.074) S.E : 0.274	1.681 (0.194) S.E : 0.210	1.664 (0.123) S.E : 0.147	1.659 (0.076) S.E : 0.093	1.705 (0.229) S.E : 0.214	1.679 (0.145) S.E : 0.151	1.665 (0.088) S.E : 0.098
ϕ_x	0.375	0.362 (0.124) S.E : 0.227	0.361 (0.083) S.E : 0.224	0.366 (0.052) S.E : 0.228	0.337 (0.148) S.E : 0.137	0.343 (0.100) S.E : 0.097	0.352 (0.063) S.E : 0.062	0.294 (0.191) S.E : 0.156	0.317 (0.129) S.E : 0.110	0.344 (0.082) S.E : 0.071
ϕ_r	0.550	0.543 (0.048) S.E : 0.068	0.545 (0.034) S.E : 0.070	0.547 (0.021) S.E : 0.077	0.525 (0.063) S.E : 0.074	0.531 (0.045) S.E : 0.052	0.538 (0.027) S.E : 0.033	0.524 (0.080) S.E : 0.086	0.532 (0.056) S.E : 0.061	0.542 (0.034) S.E : 0.039
σ_r	0.750	0.738 (0.056) S.E : 0.053	0.743 (0.038) S.E : 0.055	0.748 (0.024) S.E : 0.056	0.723 (0.087) S.E : 0.109	0.736 (0.057) S.E : 0.076	0.746 (0.034) S.E : 0.048	0.617 (0.269) S.E : 1.407	0.672 (0.173) S.E : 0.675	0.721 (0.053) S.E : 0.087
$L(\theta)$ or $J(\theta)$		-385.76	-800.93	-2015.15	0.30	0.25	0.23	7.55	5.84	4.92

3.5.2 Model selection and discussion

From the empirical investigation using MM with a large set of moments, we found that the statistical power of the model comparison test is weak and the result becomes inconclusive; in this case, we treat two models as being overlapping. Note here that we use the small sample to estimate the parameters of the NKM in which the asymptotic test of the model comparison is likely to make a Type II error; i.e. we accept the null hypothesis when the equal fit of moments is false.¹⁴

Table 3.10: Model selection using information criteria: inflation persistence

	GI (T=78)			GM (T=99)		
	ML	hybrid	forward	ML	hybrid	forward
$L(\theta)/T$	-3.96	-4.41	-4.82	-2.36	-2.69	-2.69
AIC	8.20	9.02	9.90	4.95	5.61	5.58
BIC	8.53	9.43	10.20	5.24	5.90	5.84
Ranking	1	2	3	1	3	2

Note: The backward- and forward-looking behaviors are examined using the MM estimation with auto- and cross-covariances at lag 1.

To make the formal test more elaborate, we rank the model according to the well-known information criteria. For this purpose, we suppose that the parameter estimates using MM are to be a possible minimum point in the likelihood function. Table 3.10 and 3.11 report the mean value for the log-likelihood and the model selection criterion: the cases of inflation and output persistence, respectively. Here we present MM with a small set of the moment conditions (auto- and cross-covariances at lag 1), because MM with alternative moments (auto- and cross-covariances at lag 4) yields the zero policy shock for the GI data.

According to AIC and BIC, by definition, we prefer the ML over the MM estimation with 15 moments for both GI and GM data. If the assumption of normality is not violated and the model is correctly specified, we can conclude that the ML estimation is the most efficient; this statistical inference is verified by the Monte Carlo study in the previous section. Nevertheless, the AIC and BIC of the MM estimation do not differ too much from the ML estimation; matching the auto- and cross-covariances at lag 1 can provide more or less the same efficiency as the likelihood inference. Also the statistical inference for the behavior of economic agents does not change; i.e. the hybrid variant can approximate the dynamics in inflation and output better than the model with purely forward-looking behavior when fitting the GI data: e.g. AIC = 9.02 (hybrid) < 9.90 (forward). On the other hand, the inconclusive result using the GM data shows that the pricesetting rule without indexation to past inflation (or purely forward-looking) is preferred due to its parsimonious description of the data: i.e. BIC = 5.90 (hybrid) > 5.84 (forward).

¹⁴Marmer and Otsu (2012) discuss the general optimality of comparison of misspecified models and propose a feasible approximation to the optimal test, which is more powerful than Rivers and Vuong (2002).

Table 3.11: Model selection using information criteria: output persistence

	GI (T=78)			GM (T=99)		
	ML	hybrid	forward	ML	hybrid	forward
$L(\theta)/T$	-3.97	-4.62	-7.88	-2.34	-3.09	-4.22
AIC	8.22	9.51	16.01	4.91	6.41	8.64
BIC	8.55	9.85	16.31	5.19	6.69	8.90
Ranking	1	2	3	1	2	3

Note: The backward- and forward-looking behaviors are examined using the MM estimation with auto- and cross-covariances at lag 1.

In Table 3.11, we have found essentially similar results for the output persistence; the results of the model comparison indicate that the backward-looking behavior in the IS equation is more appropriate for both GI and GM data. These exercises indicate that ML and MM have basically equivalent properties in statistical inference; they result in the same conclusion for the model comparison.¹⁵ In other words, if the chosen moment conditions are efficient, we do not find significant difference between the ML and MM estimations. Nonetheless, the formal test of HMT serves as a convenient tool for evaluating the performance of the competing models; since the data generating process is complex, we can attempt to find significant differences between two models along the lines of chosen moment conditions.

In addition, we can see from our empirical application that the moment-matching method achieves a high accuracy in taking the models to the data, but the parameter estimation becomes more uncertain than likelihood inference; i.e. wide confidence intervals. Indeed, these empirical findings show the variations in the model selection for evaluating the effect of the lagged term on the NKPC and the IS equation. Moreover, in our empirical application, if we include additional second moments in the objective function, this improves the empirical performance of the two models, but will make the comparison between them inconclusive. The take-home message from this analysis is that the power of the test can decrease with a particular set of moment conditions.

¹⁵However, remember that according to the formal test of HMT, the better fit of the hybrid variant is not significantly superior to the other model when the GM data is used. In this sense, the model comparison of HMT is more concerned with a direct comparison between the models rather than the accuracy of the approximation to the underlying data generating process.

3.6 Conclusion

This chapter considered the structural estimation of the NKM where we conducted a formal comparison of the model with purely forward-looking behavior and its hybrid variant. Especially, we examined the importance of the future expected and lagged values in the inflation and output dynamics using US data; i.e. forward- and backward-looking behavior in the NKPC and the IS equation. The models are estimated by the classical estimation methods of MM and ML. In the former, we derived the analytic moments of the auto- and cross-covariances from a linear system of the NKM; we estimated the parameters by matching the model-generated moments with their empirical counterparts. These empirical findings are compared with the ML estimation while their sensitivity to the moment conditions is also examined.

According to the estimated loss function values obtained by MM, we evaluated two competing models using the formal test of HMT when they are overlapping or one model is nested within another. The empirical results show that the inclusion of a lagged term in the NKPC and the IS equation improves the model's empirical performance. In other words, the backward-looking behavior in the model plays an important role in approximating the persistence of inflation and output. This result suggests intrinsic persistence as the main source of the inflation and output dynamics in GI. However, in GM, we cannot reject the null hypothesis at 5% level, because the model with purely forward-looking expectations and its hybrid variant have an equal fit to the data. These empirical findings are verified using the Monte Carlo experiments; we investigated the statistical efficiency of the estimators and the implications for the model selection.

We close this chapter by pointing out that (analytic) moment conditions provide a relevant information about the data generating process, which can be used to estimate structural parameters in the model; from this, we can directly compare the competing specifications in the NKM using the formal test. Moreover, if the model does not have readily available expressions for moment conditions due to its non-linear model structure, they can be replaced by an approximation based on simulations. For example, the model of De Grauwe (2011) connects the discrete choice theory to a monetary DSGE framework in which agents' belief can display endogenous waves of market optimism and pessimism. However, the non-linear variant of the DSGE model does not have a simple closed-form expression for a VAR (q) process. If this is the case, the simulated method of moments can offer an empirical analysis of the model by approximating the non-linearity in the moment conditions; e.g. see Jang and Sacht (2012) regarding simulation based inference for the non-linear group dynamics. Another example would be a DSGE model with recursive preference and stochastic volatility; i.e. see also Caldara et al. (2012) for the comparison of the solution methods. The non-linearity from recursive preferences and stochastic volatility can be simply simulated and estimated via the method of moments adopted in this chapter. We leave it to future research to empirically examine this kind of non-linear models.

Appendix

A: Choice of moments

A1: Auto- and cross-covariances at lag 1 (one quarter): 15 moment conditions

This section lists the moment conditions for the method of moment estimation. The auto- and cross-covariances at lag 1 include the following 15 moment conditions after removing double counting of the interest gap (\widehat{r}_t), the output gap (x_t), and the inflation gap ($\widehat{\pi}_t$).

- | | |
|---|---|
| 1. m_1 : $\text{Var}(\widehat{r}_t)$ | 9. m_9 : $\text{Cov}(x_t, x_{t-1})$ |
| 2. m_2 : $\text{Cov}(\widehat{r}_t, \widehat{r}_{t-1})$ | 10. m_{10} : $\text{Cov}(x_t, \widehat{\pi}_t)$ |
| 3. m_3 : $\text{Cov}(\widehat{r}_t, x_t)$ | 11. m_{11} : $\text{Cov}(x_t, \widehat{\pi}_{t-1})$ |
| 4. m_4 : $\text{Cov}(\widehat{r}_t, x_{t-1})$ | 12. m_{12} : $\text{Cov}(\widehat{\pi}_t, x_{t-1})$ |
| 5. m_5 : $\text{Cov}(\widehat{r}_t, \widehat{\pi}_t)$ | 13. m_{13} : $\text{Cov}(\widehat{\pi}_t, \widehat{r}_{t-1})$ |
| 6. m_6 : $\text{Cov}(\widehat{r}_t, \widehat{\pi}_{t-1})$ | 14. m_{14} : $\text{Var}(\widehat{\pi}_t)$ |
| 7. m_7 : $\text{Cov}(x_t, \widehat{r}_{t-1})$ | 15. m_{15} : $\text{Cov}(\widehat{\pi}_t, \widehat{\pi}_{t-1})$ |
| 8. m_8 : $\text{Var}(x_t)$ | |

A2: Auto- and cross-covariances at lag 4 (one year): 42 moment conditions

In the same vein, there are nine profiles of the sample covariance functions. Counting all the combination of three observable variables gives 42 moment conditions for the auto- and cross-covariances at lag 4. To save space, we abstract its list here by using the following notation:

$$\text{Cov}(u_t, v_{t-h}), \quad u \ \& \ v = \widehat{r}_t, x_t, \widehat{\pi}_t \quad (3.28)$$

where h denotes the lag length used in the auto- and cross-covariances ($h = 0, 1, 2, 3, 4$).

B: Reduced form of matrix and solution of the NKM

In this section we give a description of the matrix notation in Equation (3.3) and the solution procedure for the system of the NKM. The matrices of A , B , C and N with $y_t = (\widehat{\pi}_t, x_t, \widehat{r}_t)'$ are defined as follows.

$$A = \begin{bmatrix} 0 & 0 & \frac{\beta}{1+\alpha\beta} \\ 0 & \frac{1}{1+\chi} & \tau \\ 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & \kappa & -1 \\ -\tau & -1 & 0 \\ -1 & (1-\phi_r)\phi_x & (1-\phi_r)\phi_\pi \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 0 & \frac{\alpha}{1+\alpha\beta} \\ 0 & \frac{\chi}{1+\chi} & 0 \\ \phi_r & 0 & 0 \end{bmatrix}, \quad N = \begin{bmatrix} 0 & 0 & \rho_\pi \\ 0 & \rho_x & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Using Equation (3.4), we redefine the vector of observable variables y_t as terms of one-period-ahead.

$$\begin{aligned} y_{t+1} &= \Omega y_t + \Phi \nu_{t+1} \\ &= \Omega(\Omega y_{t-1} + \Phi \nu_t) + \Phi(N \nu_t + \varepsilon_{t+1}) \\ &= \Omega^2 y_{t-1} + (\Omega \Phi + \Phi N) \nu_t + \Phi \varepsilon_{t+1} \end{aligned} \quad (3.29)$$

Substitute Equations (3.29) and (3.4) into the canonical form of Equation (3.3).

$$E_t [A \Omega^2 y_{t-1} + A(\Omega \Phi + \Phi N) \nu_t + A \Phi \varepsilon_{t+1} + B \Omega y_{t-1} + B \Phi \nu_t + C y_{t-1} + \nu_t] = 0 \quad (3.30)$$

Drop the expectation and rearrange things.

$$(A \Omega^2 + B \Omega + C) y_{t-1} + (A \Omega \Phi + A \Phi N + B \Phi + I_n) \nu_t = 0 \quad (3.31)$$

where $n=3$. This implies that the following equations must hold for all y_{t-1} and ν_t .

$$\begin{aligned} A \Omega^2 + B \Omega + C &= 0 \\ (A \Omega + B) \Phi + A \Phi N + I_n &= 0 \end{aligned} \quad (3.32)$$

An iterative method can provide the solution of the matrix Ω . The matrix Φ can be obtained by using some matrix algebra; i.e. the solution of the Lyapunov equation.

C: VAR lag order selection

In our study, a VAR (q) model describes the relationship between the empirical auto- and cross-covariances of interest rate, inflation and output. We employ the model of a K -dimensional multiple times series $y_t := (y_{1t}, \dots, y_{Kt})'$ following Lütkepohl (2005):

$$y_t = \nu + A_1 y_{t-1} + \dots + A_q y_{t-q} + u_t \quad (3.33)$$

where ν is a fixed ($K \times 1$) vector of intercept, and u_t is a K -dimensional innovation process with $E(u_t) = 0$, $E(u_t u_t') = \Sigma_u$. The matrices A_i include fixed ($K \times K$) coefficients. The following lag order selection criteria are considered in Table 3.12: final prediction error (FPE), Akaike information criterion (AIC), Hannan-Quinn information criterion (HQ), Bayesian information criterion (BIC). The chosen lag order for both periods is one year (VAR (4)).

Table 3.12: VAR lag order selection criteria

Lag	GI				GM			
	FPE	AIC	HQ	BIC	FPE	AIC	HQ	BIC
0	14931.714	9.534	9.534	9.534	8926.601	9.036	9.036	9.036
1	194.525	5.309	5.302	5.466	205.437	5.554	5.558	5.699
2	106.200	4.822	4.805	5.137	112.227	4.843	4.851	5.136
3	24.202	3.462	3.435	3.936	26.806	3.505	3.515	3.945
4	1.136	0.522*	0.482*	1.156*	1.696	0.839*	0.851*	1.427*
5	1.058	0.569	0.515	1.365	1.759	0.970	0.983	1.708
6	0.944*	0.571	0.501	1.528	2.094	1.238	1.251	2.127
7	0.970	0.709	0.620	1.830	1.611	1.068	1.081	2.110
8	1.050	0.893	0.783	2.177	1.563*	1.129	1.139	2.324

Note: The star (*) indicates an optimal lag length.

D: Matrix notation

This section gives a matrix notation for the derivative of the moment conditions. This notation is used to implement the procedures for the model comparison of HMT; see appendix E. Let $m(\theta)$ be a m_n by 1 vector. The parameter vector θ has a dimension of n_θ^I . The gradient matrix $\frac{\partial m(\theta)}{\partial \theta'}$ has dimension $m_{m_n} \times n_\theta^I$. The second derivative matrix $\frac{\partial}{\partial \theta^{I'}} \text{vec} \left(\frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} \right)$ has dimension $m_{m_n} \cdot n_\theta^I \times n_\theta^I$

$$\frac{\partial m(\theta)}{\partial \theta'} = \begin{bmatrix} \frac{\partial m_1}{\partial \theta_1} & \frac{\partial m_1}{\partial \theta_2} & \cdots & \frac{\partial m_1}{\partial \theta_{n_\theta^I}} \\ \frac{\partial m_2}{\partial \theta_1} & \frac{\partial m_2}{\partial \theta_2} & \cdots & \frac{\partial m_2}{\partial \theta_{n_\theta^I}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial m_{m_n}}{\partial \theta_1} & \frac{\partial m_{m_n}}{\partial \theta_2} & \cdots & \frac{\partial m_{m_n}}{\partial \theta_{n_\theta^I}} \end{bmatrix}.$$

$$\frac{\partial}{\partial \theta^{I'}} \text{vec} \left(\frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} \right) = \begin{bmatrix} \frac{\partial m_1}{\partial \theta_1 \partial \theta_1} & \frac{\partial m_1}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial m_1}{\partial \theta_1 \partial \theta_{n_\theta^I}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial m_{m_n}}{\partial \theta_1 \partial \theta_1} & \frac{\partial m_{m_n}}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial m_{m_n}}{\partial \theta_1 \partial \theta_{n_\theta^I}} \\ \frac{\partial m_1}{\partial \theta_2 \partial \theta_1} & \frac{\partial m_1}{\partial \theta_2 \partial \theta_2} & \cdots & \frac{\partial m_1}{\partial \theta_2 \partial \theta_{n_\theta^I}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial m_{m_n}}{\partial \theta_2 \partial \theta_1} & \frac{\partial m_{m_n}}{\partial \theta_2 \partial \theta_2} & \cdots & \frac{\partial m_{m_n}}{\partial \theta_2 \partial \theta_{n_\theta^I}} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial m_1}{\partial \theta_{n_\theta^I} \partial \theta_1} & \frac{\partial m_1}{\partial \theta_{n_\theta^I} \partial \theta_2} & \cdots & \frac{\partial m_1}{\partial \theta_{n_\theta^I} \partial \theta_{n_\theta^I}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial m_{m_n}}{\partial \theta_{n_\theta^I} \partial \theta_1} & \frac{\partial m_{m_n}}{\partial \theta_{n_\theta^I} \partial \theta_2} & \cdots & \frac{\partial m_{m_n}}{\partial \theta_{n_\theta^I} \partial \theta_{n_\theta^I}} \end{bmatrix}.$$

E: Technical note on the model comparison method

This section recapitulates the equations for the model comparison method of HMT. Assume that model B is nested within model A . The quantitative goodness-of-fit of models to data is evaluated using the method of moments in section 3.1. The "full" model is tested against the "restricted" model.

Let m_T be a n_m vector of moments. $\widehat{m}(\theta)$ is the consistent estimator of m_T . The uncertainty of moment estimates is assessed by estimating a Newey-West type weighted sum of autocovariance matrices ($\widehat{\Sigma}_m$). Given the assumption of normality, we can consistently estimate the covariance matrix of moment conditions.

$$\sqrt{T}(m_T - \widehat{m}(\theta)) \xrightarrow{d} N(0, \widehat{\Sigma}_m) \quad (3.34)$$

The estimates $\widehat{\theta}^I$ are obtained at the point where a weighted objective function is minimized:

$$J(\theta^I) \equiv \min_{\theta^I \in \Theta} T \cdot \|W^{1/2}(\widehat{m}_T - m^I(\widehat{\theta}^I))\|^2, \quad I = A, B \quad (3.35)$$

$\|W^{1/2}(\widehat{m}_T - m^I(\widehat{\theta}^I))\|$ is defined as $\sqrt{(\widehat{m}_T - m^I(\widehat{\theta}^I))'W(\widehat{m}_T - m^I(\widehat{\theta}^I))}$. The weight matrix W is set to the diagonal components of $1/\widehat{\Sigma}_{m,ii}$ ($ii = 1, \dots, n_m$). The quasi-likelihood ratio test statistic is constructed as the difference in fits between two models:

$$\text{QLR}(\widehat{\theta}^B, \widehat{\theta}^A) = J^B(\widehat{\theta}^B) - J^A(\widehat{\theta}^A) \quad (3.36)$$

J^I ($I = A, B$) is a minimum value of the objective function given parameter estimates from Equation (3.35). It is assumed that the chosen moment functions in the models are twice continuously differentiable in neighborhoods of $\theta^I \subset \Theta^{n_\theta^I}$. Further, the matrix F and M are non-singular in neighborhoods of θ .¹⁶:

$$F^I = \frac{\partial m^I(\theta^I)'}{\partial \theta^I} W \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} - M^I \quad (3.37)$$

$$M^I = (E_I \otimes (\widehat{m}_T - m^I(\theta^I))'W) \frac{\partial}{\partial \theta^{I'}} \text{vec}\left(\frac{\partial m^I(\theta^I)}{\partial \theta^{I'}}\right), \quad I = A, B \quad (3.38)$$

E_I is the identity matrix of which dimension is $n_\theta^I \times n_\theta^I$. Note here that the dimen-

¹⁶We use the built-in procedures `gradp` and `hessp` in the GAUSS software package. The optimal step size for the gradient vector and the Hessian matrix is carefully adjusted, because difference approximations is likely to be imprecise provided that the first derivative is small. See Gill et al. (1981, Ch.4, pp. 127-133) for the choice of the finite-difference interval.

sions of the matrices $\frac{\partial m^I(\theta^I)}{\partial \theta^{I'}}$ and $\frac{\partial}{\partial \theta^{I'}} \text{vec}\left(\frac{\partial m^I(\theta^I)}{\partial \theta^{I'}}\right)$ are $n_m \times n_\theta^I$ and $n_m \cdot n_\theta^I \times n_\theta^I$. The dimension of F^I and M^I are n_θ^I by n_θ^I .

The theorem 3.1 in HMT states that the quasi-likelihood ratio test QLR converges in distribution to Equation (3.18). The n_θ^I by n_θ^I matrix V^I is defined as $V^I = V_1^I - V_2^I - V_3^I$ with $I = A, B$:

$$\begin{aligned} V_1^I &= \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} (F^{I'})^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} W \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} (F^I)^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} \\ V_2^I &= \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} ((F^{I'})^{-1} + (F^I)^{-1}) \frac{\partial m^I(\theta^I)'}{\partial \theta^I} \\ V_3^I &= \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} (F^{I'})^{-1} (M^{I'} + M^I) (F^I)^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} \end{aligned}$$

However, it is sometimes observed that the estimated $\hat{V}_B - \hat{V}_A$ is not a positive-definite matrix where some negative values are drawn in simulations. We should not discard the negative values of the test distribution when making statistical inference for the model comparison. The hypothesis test is assessed by critical values at the 1% and 5% confidence level (Q_{99}, Q_{95}) from the simulated asymptotic test distribution. When one model is nested within another, one rejects the null hypothesis at 5% level that two models are equivalent if $\text{QLR}(\hat{\theta}_A, \hat{\theta}_B) > Q_{95}$.

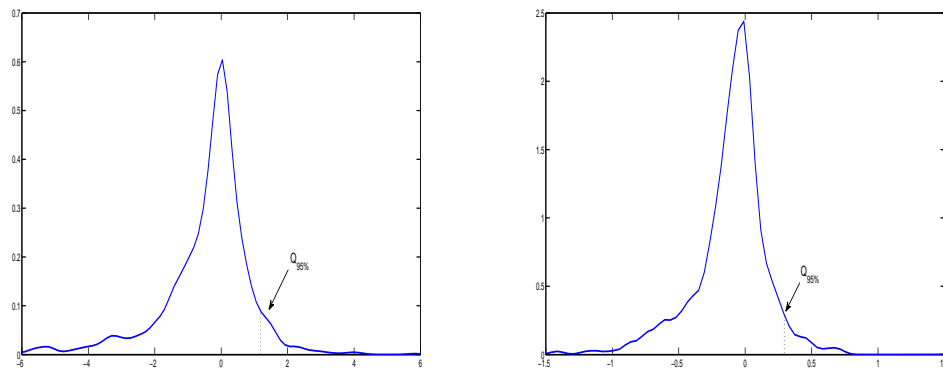
F: Simulated QLR distribution for model comparison**F1: Auto- and cross-covariances at lag 1: 15 moment conditions**

Figure 3.3: Test distribution for inflation persistence: GI (left) and GM (right)

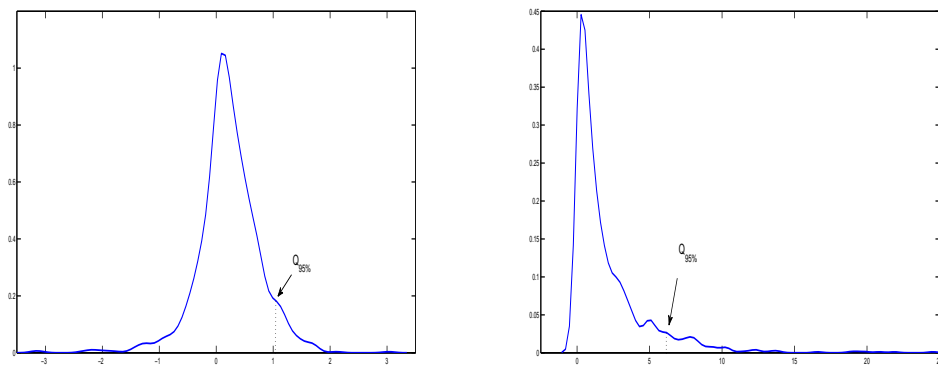


Figure 3.4: Test distribution for output persistence: GI (left) and GM (right)

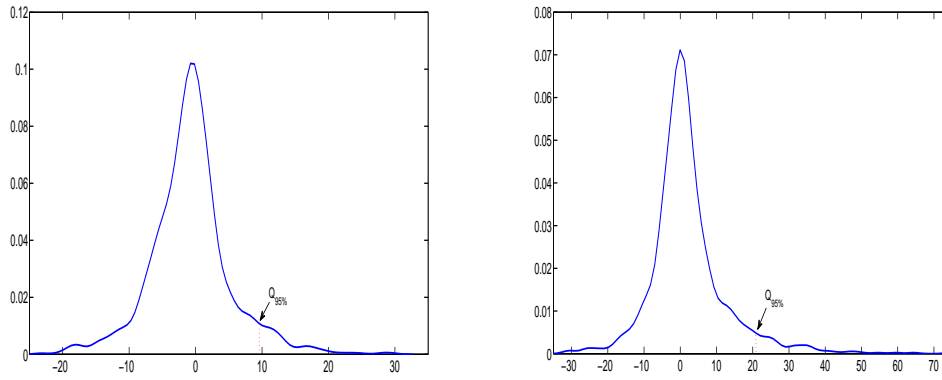
F2: Auto- and cross-covariances at lag 4: 42 moment conditions

Figure 3.5: Test distribution for inflation persistence: GI (left) and GM (right)

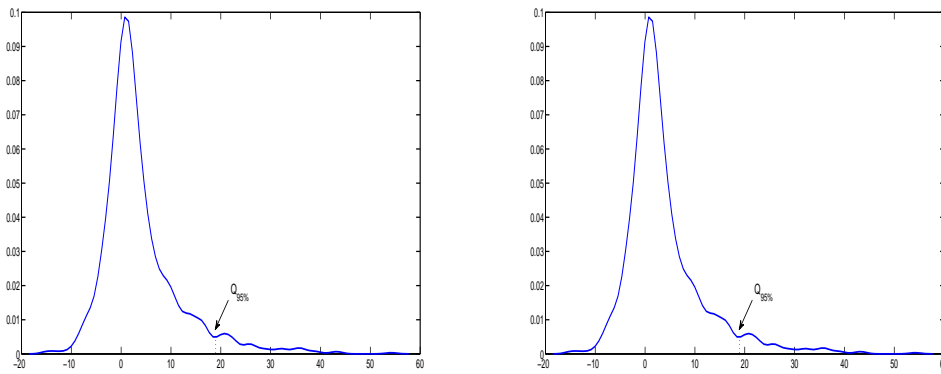


Figure 3.6: Test distribution for output persistence: GI (left) and GM (right)

G: The Monte Carlo result of the misspecified case

Table 3.13: Monte Carlo results on the MM and ML estimates of the misspecified model, (): root mean square error, S.E : mean of standard errors

θ^0	ML			MM with 15 moments			MM with 42 moments			
	T = 100	T = 200	T = 500	T = 100	T = 200	T = 500	T = 100	T = 200	T = 500	
κ	0.050	0.096 (0.186)	0.089 (0.212)	0.077 (0.031)	0.176 (0.140)	0.168 (0.125)	0.163 (0.118)	0.205 (0.175)	0.191 (0.152)	0.182 (0.136)
ρ_π	0.000	0.616 (0.621)	0.618 (0.620)	0.617 (0.618)	0.632 (0.635)	0.646 (0.647)	0.653 (0.654)	0.598 (0.604)	0.614 (0.617)	0.623 (0.624)
σ_π	0.675	0.491 (0.293)	0.487 (0.330)	0.474 (0.205)	0.560 (0.151)	0.543 (0.150)	0.531 (0.654)	0.661 (0.164)	0.633 (0.127)	0.612 (0.098)
χ	1.000	0.921 (0.132)	0.938 (0.100)	0.955 (0.066)	0.981 (0.053)	0.994 (0.020)	0.999 (0.015)	0.970 (0.083)	0.986 (0.047)	0.997 (0.014)
τ	0.090	0.085 (0.032)	0.085 (0.024)	0.085 (0.015)	0.089 (0.029)	0.086 (0.021)	0.084 (0.014)	0.088 (0.035)	0.083 (0.024)	0.080 (0.017)
σ_x	0.700	0.688 (0.064)	0.691 (0.046)	0.694 (0.026)	0.637 (0.123)	0.636 (0.103)	0.636 (0.082)	0.654 (0.132)	0.644 (0.106)	0.639 (0.083)
ϕ_π	1.650	1.667 (0.182)	1.657 (0.118)	1.657 (0.075)	1.691 (0.182)	1.681 (0.117)	1.679 (0.075)	1.848 (0.291)	1.783 (0.203)	1.775 (0.156)
ϕ_x	0.375	0.352 (0.127)	0.352 (0.085)	0.356 (0.054)	0.227 (0.211)	0.227 (0.203)	0.226 (0.164)	0.315 (0.282)	0.238 (0.197)	0.237 (0.166)
ϕ_r	0.550	0.540 (0.049)	0.541 (0.035)	0.356 (0.054)	0.488 (0.086)	0.487 (0.077)	0.489 (0.067)	0.527 (0.070)	0.524 (0.053)	0.525 (0.038)
σ_r	0.750	0.738 (0.056)	0.743 (0.039)	0.748 (0.024)	0.733 (0.101)	0.744 (0.069)	0.756 (0.043)	0.597 (0.313)	0.616 (0.244)	0.649 (0.164)
$L(\theta)$ or $J(\theta)$		-398.38	-805.68	-2026.45	2.36	3.46	6.95	24.22	29.36	49.73

Note: The misspecified model does not include the parameter α in the NKPC. To save space, we do not report the asymptotic standard errors for the parameter estimates, because these are not qualitatively different from the correctly specified case.

Identification of Animal Spirits in a Bounded Rationality Model

In this chapter we empirically examine a heterogeneous bounded rationality version of a hybrid New-Keynesian model. The model is estimated via the simulated method of moments using Euro Area data from 1975Q1 to 2009Q4. It is generally assumed that agents' beliefs display waves of optimism and pessimism - so called animal spirits - on future movements of the output and inflation gap. Our main empirical findings show that a bounded rationality model with cognitive limitation provides a reasonable fit to auto- and cross-covariances of the data. This result is mainly driven by a high degree of intrinsic persistence in the output and inflation gap due to the impact of animal spirits on economic dynamics. Further, over the whole time interval the agents had expected moderate deviations of the future output gap from its steady state value with low uncertainty. Finally, we find strong evidence for an autoregressive expectation formation process regarding the inflation gap.

4.1 Introduction

Rational expectations are a flexible and natural way of modeling market behavior in dynamic stochastic general equilibrium (DSGE) models, which are widely used by macroeconomists. Since the DSGE approach disposes a convenient analytical tractability under the assumption of rational expectations, this modeling framework serves as an efficient toolbox for analyzing monetary and fiscal policy measures. As Selten (2001) states, however, "modern mainstream economic theory is largely based on an unrealistic picture of human decision theory" since evidence from experimental studies supports information processing with limited cognitive ability of agents rather than perfect information (see Hommes (2011) among others). Indeed, a plethora of studies have been done on alternative forms of information processing mechanisms in macroeconomics; see e.g. the literature on learning (Evans and Honkapohja (2001)), rational inattention (Sims (2003)), sticky information (Mankiw and Reis (2002)) or bounded rationality in general (Sargent (1994) and Kahneman (2003)). Camerer (1998) also offers an informative overview of the discussion on this topic in economics.

For the most part of the behavioral research, we can treat the realization of economic decisions as being a complex and interactive process between different types of agents. Keynes (1936) already attributed significant irrationality to human

nature and discussed the impacts of waves of optimism and pessimism - so called *animal spirits* - on economic outcome. According to Akerlof and Shiller (2009), the emotional states are reflected in economic behavior - see also Franke (forthcoming) for his extensive discussion about market behavior and how expectation formation should be treated in macroeconomic models.

In this chapter we attempt to empirically examine the hypothesis that the behavioral heterogeneity will have a macroscopic impact on the economy. The point of view taken here is that a behavioral model can provide a conceptual framework for a cognitive ability as well as a substantial degree of inertia in the DSGE models. According to De Grauwe (2011), if agents are known to be either optimists or pessimists, their ability (or better: limitation) to form their expectations affects economic conditions, i.e. movements in employment, the output gap and inflation, more appropriately than standard rational expectation models. Indeed, it is shown in the expectation formation process under bounded rationality that we can explicitly model animal spirits by applying discrete choice theory on group behavior. Then the behavior of optimists and pessimists is considered to be a by-product of the switching mechanisms based on the performance measure from agents' expectations (see also e.g. Westerhoff (2008) as well as Lengnick and Wohltmann (forthcoming) among others).

To the best of our knowledge, however, an empirical evaluation of a bounded rationality model of this type discussed above is missing in the literature so far. We fill the existing gap between the use of the models and their empirical evaluations in the literature by measuring the effects of psychological behavior on the economy under consideration of animal spirits. We show that the moment-based estimation (Franke et al. (2011)) can be easily used to estimate a small-scale DSGE model. Mainly, similarities and dissimilarities between two polar cases of expectation formation processes will be examined: while the underlying model structure is identical to a standard three-equations New-Keynesian model (NKM), we also allow both for rational expectations and for endogenously-formed expectations using the behavioral specification by De Grauwe (2011). In particular, we study his behavioral economic framework and provide an empirical investigation of bounded rationality on economic dynamics in the Euro Area from 1975Q1 to 2009Q4. Accordingly, an important aspect of this chapter is to test the bounded rationality hypothesis in order to offer reliable parameter values that can be used for calibration in more realistic-grounded future work, e.g. studying monetary and fiscal policy analysis in a DSGE model without the assumption of rational expectations.

In our empirical application, we show that the NKM with rational expectations or bounded rationality can generate auto- and cross-covariances of the output gap, the inflation gap and the interest gap, which can mimic real data well. A quadratic objective function is used in the estimation to measure the distance between the model-generated and empirical moments. As the usual procedure of the method of moments, the global minimum of the objective function provides consistent parameter estimates of the model. Then we evaluate the goodness-of-fit of the model to the data from the value of the quadratic object function; i.e. the lower this value, the

better the fit of the model-generated moments to their empirical counterparts. The empirical application using the method of moment approach stays in line with the work of Franke et al. (2011), who estimate a similar version of the NKM presented here for two sub-samples, i.e. the Great Inflation and Great Moderation period in the US. They come to the conclusion that inflation dynamics are primarily driven by intrinsic rather than extrinsic persistence - which is the total opposite of the results when applying Bayesian estimation. This is reflected by a high degree of price indexation and a low degree of persistence in the assumed AR(1) cost-push shock. In general, this kind of estimation technique is closely related to the approaches of indirect inference with the difference that in our case the structural form of a DSGE model is used instead of an auxiliary model like a SVAR (cf. Smith (1993) and Christiano et al. (2005) among others).

Main findings can be summarized as follows. First, over the whole time interval the agents had expected moderate deviations of the future output gap from its steady state value with low uncertainty. Second, we find strong evidence for an autoregressive expectation formation process regarding the inflation gap, which is in line with the scientific consensus among experimental economists (Roos and Schmidt (2012)).

The remainder of the chapter is structured as follows. Section 2 introduces a small-scale NKM and discusses two model specifications, i.e. one with rational expectations and the other under consideration of the animal spirits. The estimation methodology is presented in section 3. Section 4 then estimates two versions of the model by the moment-based estimation and discusses their empirical results. Afterwards, the properties of the moment-based procedure for estimation are examined through a Monte Carlo study and a sensitivity analysis in section 5. Finally, section 6 concludes. The appendix collects all relevant technical details.

4.2 The model: rational expectations vs. bounded rationality

The New-Keynesian three-equations model reads as follows:

$$y_t = \frac{1}{1+\chi} \tilde{E}_t^j y_{t+1} + \frac{\chi}{1+\chi} y_{t-1} - \tau(\hat{r}_t - \tilde{E}_t^j \hat{\pi}_{t+1}) + \varepsilon_{y,t} \quad (4.1)$$

$$\hat{\pi}_t = \frac{\nu}{1+\alpha\nu} \tilde{E}_t^j \hat{\pi}_{t+1} + \frac{\alpha}{1+\alpha\nu} \hat{\pi}_{t-1} + \kappa y_t + \varepsilon_{\hat{\pi},t} \quad (4.2)$$

$$\hat{r}_t = \phi_{\hat{r}}(\phi_{\hat{\pi}} \hat{\pi}_t + \phi_y y_t) + (1 - \phi_{\hat{r}}) \hat{r}_{t-1} + \varepsilon_{\hat{r},t} \quad (4.3)$$

where the superscript $j = \{\text{RE}, \text{BR}\}$ refers to the rational expectation (RE) and the bounded rationality (BR) model, which we describe below. The corresponding expectations operator is \tilde{E}_t^j , which has to be specified for both models. It goes without saying that all variables are given in quarterly magnitudes. Equation (4.1) describes a hybrid dynamic IS curve and results from the standard utility maximization approach of a representative household. Here the current output gap depends

negatively on the real interest rate, i.e. it is stemming from intertemporal optimization of consumption and saving resulting in consumption smoothing. The parameter $\tau \geq 0$ denotes the inverse intertemporal elasticity of substitution. Equation (4.2) is known as the hybrid New-Keynesian Phillips Curve (NKPC) where the output gap (y_t) is the driving force of inflation due to monopolistic competition and the Calvo price-setting scheme. The slope of the Phillips Curve is given by the parameter $\kappa \geq 0$. The parameter ν denotes the discount factor ($0 < \nu < 1$). According to the Taylor rule with interest rate smoothing (4.3), the nominal interest gap is a predetermined variable while the monetary authority reacts directly to movements in the output ($\phi_y \geq 0$) and inflation ($\phi_{\hat{\pi}} \geq 0$) gap. We account for intrinsic persistence in this stylized version of the well-known Smets and Wouters (2003, 2005 and 2007) model due to the assumption of backward-looking behavior indicated by the parameters for habit formation χ , price indexation α and interest rate smoothing $\phi_{\hat{r}}$, respectively ($0 \leq \chi \leq 1$, $0 \leq \alpha \leq 1$, $0 \leq \phi_{\hat{r}} \leq 1$). We assume that the exogenous driving forces in the model variables follow idiosyncratic shocks $\varepsilon_{z,t}$, which are drawn from multivariate normal distributions around mean zero and variance σ_z^2 with variables $z = \{y, \hat{\pi}, \hat{r}\}$.

Note here that we consider the gaps instead of the levels and therefore account explicitly for a time-varying trend in inflation and the natural rate of interest. The corresponding gaps are simply given by taking the difference of the actual value for output, inflation and the interest rate from their trends (i.e. time-varying steady state values) respectively where the latter is computed by applying the Hodrick-Prescott filter with a standard value of the corresponding smoothing parameter of 1600. Accordingly, the set of equations is used to describe the dynamics in the output gap y_t , the inflation gap $\hat{\pi}_t$ and the nominal interest rate gap \hat{r}_t , where \hat{x}_t with $x = \{\pi, r\}$ denotes the deviations in both variables from the time-varying trend explicitly.

The results of many studies show that assuming a constant trend, like a zero-inflation steady state, leads to misleading results. For example, Ascari and Ropele (2009) observe that the dynamic properties (i.e. mainly the stability of the system) depend on the variation in trend inflation. Cogley and Sbordone (2008) also provide evidence for the explanation of inflation persistence by considering a time-varying trend in inflation. In the same vein, we can abandon the assumption of a constant natural rate of interest as being empirically unrealistic. In this chapter, we follow the empirical approaches proposed by Cogley et al. (2010), Castelnuovo (2010), Franke et al. (2011) among others, who also consider gap specifications for inflation (and the nominal interest rate). Furthermore, inflation and money growth are likely to be non-stationary in the Euro Area data. If that is the case, the estimation methodology such as the method of moments approach presented here (or the generalized method of moments in general) will lead to biased estimates.¹ Taken this into account, in the current study we consider the gaps rather than the levels in order to

¹See also Russel and Banerjee (2008) as well as Aussenmacher-Wesche and Gerlach (2008) among others for methodological issues related to non-stationary inflation in the US and the Euro Area.

ensure the stationary of the times series.

To make the description of the expectation formation processes more explicit, first we examine two polar cases in the theoretical model framework of the NKM. First, under rational expectations, the forward-looking terms, which are the expectations of the output gap and inflation gap at time $t+1$ in equations (4.1) and (4.2), are just given by

$$\tilde{E}_t^{RE} y_{t+1} = E_t y_{t+1} \quad (4.4)$$

$$\tilde{E}_t^{RE} \hat{\pi}_{t+1} = E_t \hat{\pi}_{t+1} \quad (4.5)$$

where E_t denotes the expectations operator conditional on information given at time t . Second, as regards the other specification, we depart from rational expectations by considering a behavioral model of De Grauwe (2011). It is generally assumed that agents will be either *optimists* or *pessimists* (in the following indicated by the superscripts O and P , respectively) who form expectations based on their beliefs regarding movements in the future output gap:

$$E_t^O y_{t+1} = d_t \quad (4.6)$$

$$E_t^P y_{t+1} = -d_t \quad (4.7)$$

where

$$d_t = \frac{1}{2} \cdot [\beta + \delta \sigma(y_t)] \quad (4.8)$$

"can be interpreted as the divergence in beliefs among agents about the output gap" (De Grauwe (2011, p. 427)). In contrast to the RE model, both types of agents are uncertain about the future dynamics of the output gap and therefore predict a fixed value of y_{t+1} denoted by $\beta \geq 0$. We can interpret the latter as the *predicted subjective mean value* of y_t . However, this kind of subjective forecast is generally biased and therefore depends on the volatility in the output gap; i.e. given by the unconditional standard deviation $\sigma(y_t) \geq 0$. In this respect, the parameter $\delta \geq 0$ measures the *degree of divergence* in the movement of economic activity. Note that due to the symmetry in the divergence in beliefs, optimists expect that the output gap will differ positively from the steady state value (which for consistency is set to zero) while pessimists will expect a negative deviation by the same amount. The value of δ remains the same across both types of agents.

The expression for the market forecast regarding the output gap in the bounded rationality model is given by

$$\tilde{E}_t^{BR} y_{t+1} = \alpha_{y,t}^O \cdot E_t^O y_{t+1} + \alpha_{y,t}^P \cdot E_t^P y_{t+1} = (\alpha_{y,t}^O - \alpha_{y,t}^P) \cdot d_t \quad (4.9)$$

where $\alpha_y^O + \alpha_y^P = 1$. A specific forecasting rule chosen by agents, i.e. (4.6) or (4.7), is indicated by the probability of $\alpha_{y,t}^O$ and $\alpha_{y,t}^P$, respectively. In particular, α_y^O (or α_y^P) can also be interpreted as the probability being an optimist (or pessimist). In the following, we show explicitly how these probabilities are computed. Indeed, the selection of the forecasting rules (4.6) or (4.7) depends on the forecast performances

94 4. Identification of Animal Spirits in a Bounded Rationality Model

of optimists and pessimists U_t^k given by the mean squared forecasting error, which can be simply updated in every period as

$$U_t^k = \rho U_{t-1}^k - (1 - \rho)(E_{t-1}^k y_t - y_t)^2 \quad (4.10)$$

where $k = O, P$ and the parameter ρ denotes the measure of the memory of agents ($0 \leq \rho \leq 1$). Here $\rho = 0$ means that agents have no memory of past observations while $\rho = 1$ means that they have infinite memory instead. By applying discrete choice theory under consideration of the forecast performances, agents revise their expectations in which different performance measures will be utilized for $\alpha_{y,t}^O$ and $\alpha_{y,t}^P$.²

$$\alpha_{y,t}^O = \frac{\exp(\gamma U_t^O)}{\exp(\gamma U_t^O) + \exp(\gamma U_t^P)} \quad (4.11)$$

$$\alpha_{y,t}^P = \frac{\exp(\gamma U_t^P)}{\exp(\gamma U_t^O) + \exp(\gamma U_t^P)} = 1 - \alpha_{y,t}^O \quad (4.12)$$

where the parameter $\gamma \geq 0$ denotes the intensity of choice: if $\gamma = 0$, the self-selecting mechanism is purely stochastic ($\alpha_{y,t}^O = \alpha_{y,t}^P = 1/2$), whereas if $\gamma = \infty$, it is fully deterministic ($\alpha_{y,t}^O = 0$, $\alpha_{y,t}^P = 1$ or vice versa; see De Grauwe (2011), p. 429). For clarification, if $\gamma = 0$ agents are indifferent in being optimist or pessimist while if $\gamma = \infty$ their expectation formation process is independent of their emotional state, i.e. they react quite sensitively to infinitesimal changes in their forecast performances.

We explain this revision process as follows. Given the past value of the forecast performance (U_{t-1}^k), the lower the difference between the expected value of the output gap (taken from the previous period, i.e. $E_{t-1}^k y_t = |d_{t-1}|$) and its realization in period t , the higher the corresponding forecast performance U_t^k will be. In other words, if e.g. the optimists predict future movements in y_t more accurately compared to the pessimists, then this results in $U_t^O > U_t^P$. Hence, the pessimists revise their expectations by switching to the forecasting rule used by the optimists, which we can express as $E_t^O y_{t+1} = d_t$. Finally, this forecasting rule becomes dominant and the share of pessimistic group in the market decreases. Based on the equations (4.10) to (4.12), we can rationalize equation (4.9) by using simple substitution. This results in a higher degree of volatility in the expectation formation process regarding the output gap when compared to the outcome in the RE model (we refer to section 4.2 for a clarification).

The same logic can be applied for the inflation gap expectations. Following the behavioral heterogeneity approach proposed by De Grauwe (2011, pp. 436), we assume that agents will be either so called *inflation targeters* (*tar*) or *extrapolators* (*ext*).³ In the former case, the central bank anchors expectations by announcing

²See also Westerhoff (2008, p. 199) and Lengnick and Wohltmann (forthcoming) among others for an application of discrete choice theory to models in finance and macroeconomics.

³This concept of behavioral heterogeneity has been widely used in financial market models, see e.g. Chiarella and He (2002) as well as Hommes (2006) among others.

a target for the inflation gap $\bar{\pi}$. From the view of the inflation targeters, we consider this pre-commitment strategy to be fully credible. Hence the corresponding forecasting rule becomes

$$E_t^{tar} \hat{\pi}_{t+1} = \bar{\pi} \quad (4.13)$$

where we assume $\bar{\pi} = 0$.⁴ The extrapolators form their expectations in a static way and will expect that the future value of the inflation gap equals simply its past value, i.e.

$$E_t^{ext} \hat{\pi}_{t+1} = \hat{\pi}_{t-1}. \quad (4.14)$$

This results in the market forecast for the inflation gap similar to (4.9):

$$\tilde{E}_t^{BR} \hat{\pi}_{t+1} = \alpha_{\hat{\pi},t}^{tar} E_t^{tar} \hat{\pi}_{t+1} + \alpha_{\hat{\pi},t}^{ext} E_t^{ext} \hat{\pi}_{t+1} = \alpha_{\hat{\pi},t}^{tar} \bar{\pi} + \alpha_{\hat{\pi},t}^{ext} \hat{\pi}_{t-1}. \quad (4.15)$$

The forecast performances of inflation targeters and extrapolators are given by the mean squared forecasting error written as

$$U_t^s = \rho U_{t-1}^s - (1 - \rho)(E_{t-1}^s \hat{\pi}_t - \hat{\pi}_t)^2 \quad (4.16)$$

where $s = (tar, ext)$, and finally we may write:

$$\alpha_{\hat{\pi},t}^{tar} = \frac{\exp(\gamma U_t^{tar})}{\exp(\gamma U_t^{tar}) + \exp(\gamma U_t^{ext})} \quad (4.17)$$

$$\alpha_{\hat{\pi},t}^{ext} = \frac{\exp(\gamma U_t^{ext})}{\exp(\gamma U_t^{tar}) + \exp(\gamma U_t^{ext})} = 1 - \alpha_{\hat{\pi},t}^{tar}. \quad (4.18)$$

Here $\alpha_{\hat{\pi},t}^{tar}$ denotes the probability to be an inflation targeter, which is the case if the forecast performance using the announced inflation gap target is superior to the extrapolation of the inflation gap expectations and vice versa. Note here that the memory (ρ) as well as the intensive of choice parameter (γ) do not differ across the expectation formation processes in terms of the output and inflation gap. In the end, the bounded rationality model turns out to be purely backward-looking (cf. equations (4.10) and (4.16)) while the forward- and backward-looking behavior is contained in the rational expectation model. The solution to both systems can be computed by backward-induction and the method of undetermined coefficients respectively, which are shown in appendix A.

Finally, one may argue that the presented model is not suitable for e.g. policy analysis since it is not based completely on micro-foundations. In particular, the expectation mechanisms are imposed ex post on a system of structural equations which themselves have been derived from maximizing behavior under the assumption of rational expectations. However, evidence from experimental economics can help us to motivate the assumption of the divergence in beliefs (reflects guessing) and the existence of the extrapolators (which might be seen as pattern-based time-series

⁴In this respect (based on a optimal monetary policy strategy), an inflation *gap* target of zero percent implies that the European Central Bank seeks to minimize the deviation of its (realized) target *rate* of inflation from the corresponding time-varying steady state value, where in the optimum this deviation should be zero.

forecasting) done by De Grauwe (2011) and adopted in our study. Roos and Schmidt (2012) find evidence for a backward-looking behavior in forming expectations by non-professionals in economic theory and policy. In their experimental study, they show that the projections of the future realizations in the output gap and inflation are based either on historical patterns of the time series or - in the case of no available information - on simple guessing.

From a theoretical point of view, Branch and McGough (2009) introduce heterogeneous expectations into a New Keynesian framework where the forward looking expressions in the IS curve and NKPC are convex combinations of backward- and forward-looking behavior. The authors show that a micro-founded NKM under bounded rationality can be derived if specific axioms are considered within the optimizing behavior of households and firms. These axioms ensure the ability of agents to forecast future realization of the output gap and inflation on the micro level as well as the aggregation of this behavior on the macro level. In comparison, De Grauwe (2010) allows for a switching mechanism based on discrete choice theory. It is an open question if the latter fulfills the axioms imposed by Branch and McGough (2009) which may help to overcome the (neglected) problem of mis-specification. To sum up, there is no doubt that an extensive elaboration on the microfoundation of expectations formation is needed, even though up to now it is a fact that among neuroscientists the evidence on information processing in the human brain is ambiguous.

4.3 The estimation methodology

Over the last decade the Bayesian estimation became the most popular method for the estimation of DSGE models while pushing classical estimation methods aside such as the generalized method of moments and the pure maximum likelihood approach. Indeed, the Bayesian approach certainly has the advantage over the others: on the one hand, the distributions of the parameters in a system of equations framework can be easily computed from user friendly software like e.g. Dynare. On the other hand, however, there are two major disadvantages when we apply Bayesian techniques to our empirical study.

First, the Bayesian approach to the DSGE model requires the choice of appropriate prior distributions associated with the underlying economic interpretation of the structural parameters. It is still an open question what criteria are suited best in order to identify the most accurate prior information. For instance, Lombardi and Nicoletti (2011) discuss the sensitivity of posterior estimation results to the choice of different expressions of the prior knowledge; Del Negro and Schorfheide (2008) also provide an explicit method for constructing prior distributions based on the beliefs regarding macroeconomic indicators. However, so far the existing knowledge by neuroscientists does not allow for pinning down a general micro-founded model on information processing (De Grauwe (2011)). In addition, the Bayesian estimation must be designed to cope with the shape of the prior distribution, which is

often unspecified, i.e. 'uninformative' priors; as a result, the estimated posterior becomes quite similar to the prior distribution. In this respect, the Bayesian analysis is not a panacea for the BR model, since prior information is not available at least for the behavioral parameters β , δ and ρ . Second, due to the fact that a logistic function is applied on the parameters of the BR model (as a result of applying the discrete choice theory), a researcher must use a Bayesian full-information analysis such as a particle filter. Especially, as long as this filter method is applied for evaluating the likelihood function, the estimation can be subjected to e.g. an increase in approximation errors of the non-linear model (DeJong and Dave (2007), Chap. 11).

To avoid these disadvantages of the Bayesian approach, in this chapter we seek to match the model-generated autocovariances of the interest gap, the output gap and inflation gap with their empirical counterparts. We minimize the distance between these model-generated and empirical moments under consideration of a quadratic function, which summarizes the characteristics of empirical data. This method is called simply moment matching (cf. Franke et al. (2011)). Main advantage of this econometric method is that we can check transparently the goodness-of-fit of the model to data, since the empirical comparison (graphically) between the match of the estimated and simulated autocovariances is direct.

The method of moment approach comprises distributional properties of empirical data X_t , $t = 1, \dots, T$. The sample covariance matrix at lag k is defined by

$$m_t(k) = \frac{1}{T} \sum_{t=1}^{T-k} (X_t - \bar{X})(X_{t+k} - \bar{X})' \quad (4.19)$$

where $\bar{X} = (1/T) \sum_{t=1}^T X_t$ is the vector of the sample mean. The sample average of discrepancy between the model-generated and the empirical moments is denoted as

$$g(\theta; X_t) \equiv m_t^* - m_t \quad (4.20)$$

where m_t^* is the empirical moment function and m_t the model-generated moment function (cf. equation (4.19)). θ is a $l \times 1$ vector of unknown structural parameters with a parameter space Θ . Given that the length of the business cycles lies between (roughly) one and eight years in the Euro Area. A reasonable compromise is a length of two years. Therefore we will use auto- and cross-covariances of the interest rate gap, the output gap and the inflation gap at a lag k , where $k = 0, \dots, 8$. We have a p -dimensional vector of moment conditions ($p = 78$) by avoiding double counting at the zero lags in the cross relationships.⁵

We obtain the parameter estimates from the following quadratic objective function (or loss function) as a result of the minimization process:

⁵The Delta method is used to compute the confidence bands in the auto- and cross-covariance moment estimation (see appendix B for details).

$$Q(\theta) = \arg \min_{\theta \in \Theta} g(\theta; X_t)' \widehat{W} g(\theta; X_t) \quad (4.21)$$

with the weight matrix \widehat{W} estimated consistently in several ways (see Andrews (1991)). Here we use the heteroscedasticity and autocorrelation consistent (HAC) covariance matrix estimator suggested by Newey and West (1987). The kernel estimator has the following general form with the covariance matrix of the appropriately standardized moment conditions:

$$\widehat{\Gamma}_T(j) = \frac{1}{T} \sum_{t=j+1}^T (m_t - \bar{m})(m_{t-j} - \bar{m})' \quad (4.22)$$

where \bar{m} once again denotes the sample mean. Following an automatic selection for the lag length, we use a popular choice of $j \sim T^{1/3}$ leading to $j = 5$ when estimating the covariance matrix (Newey and West (1994)):

$$\widehat{\Omega}_{NW} = \widehat{\Gamma}_T(0) + \sum_{j=1}^5 \left(\widehat{\Gamma}_T(j) + \widehat{\Gamma}_T(j)' \right). \quad (4.23)$$

The weight matrix \widehat{W} is computed from the inverse of the estimated covariance matrix. However, a high correlation between the moment conditions that we consider makes the estimated covariance matrix nearly singular. In addition, the moment conditions and the elements of the weight matrix are highly correlated when the small sample size is used (Altonji and Segal (1996)). Therefore, we use the diagonal matrix entries as the weighting scheme, i.e. we ignore the off-diagonal components of the matrix $\widehat{W} = \widehat{\Omega}_{ii}^{-1}$. The estimated confidence bands, then, become wider since the sandwich elements in the covariance of parameter estimates cannot cancel out with this weighting scheme (see also Anatolyev and Gospodinov (2011)).

Under certain regularity conditions, one can derive the following asymptotic distribution of the method of moments estimation of the parameters:

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) \sim N(0, \Lambda) \quad (4.24)$$

where $\Lambda = [(DWD')^{-1}]D'W\Omega WD[(DWD')^{-1}]'$, and D is the gradient vector of moment functions evaluated around the point estimates:

$$\widehat{D} = \left. \frac{\partial m(\theta; X_T)}{\partial \theta} \right|_{\theta = \widehat{\theta}_T}. \quad (4.25)$$

Under RE, we can obtain the simple analytic moment conditions of the model. However, for the BR model, the analytic expressions for the moment conditions are not readily available due to the non-linear discrete choice framework. To circumvent this problem, we use the simulated method of moments to estimate the behavioral parameters in the BR model. The simulated method of moments is particularly suited to a situation where the model is easily simulated by replacing theoretical moments. Then the model-generated moments in Equation (5.21) are replaced by their simulated counterparts:

$$m_t = \frac{1}{S \cdot T} \sum_{t=1}^{S \cdot T} \tilde{m}_t. \quad (4.26)$$

We can simulate the data from the model and compute the moment conditions (\tilde{m}_t) in order to approximate the theoretical moments (m_t). The simulation size is denoted by S . The asymptotic normality of the simulated method of moments holds under certain regularity conditions (Duffie and Singleton (1993), Lee and Ingram (1991)):

$$\sqrt{T}(\hat{\theta}_{SMM} - \theta_0) \sim N(0, \Lambda_{SMM}) \quad (4.27)$$

where $\Lambda_{SMM} = (B'WB)^{-1}B'W(1 + 1/S)\Omega WB(B'WB)^{-1}$, i.e. a covariance matrix of the SMM estimates. A gradient vector of the moment function is defined as $B \equiv E\left[\frac{\partial m_t}{\partial \theta}\bigg|_{\theta=\hat{\theta}_1}\right]$. Since the covariance matrix becomes less accurate than the estimation where the analytic moments are used, the model estimation is now subjected to simulation errors. To reduce the simulation error, we set the simulation size to a reasonably large value 100.

Finally, we use the J test to evaluate compatibilities of the moment conditions:

$$J \equiv T \cdot Q(\hat{\theta}) \xrightarrow{d} \chi_{p-l}^2 \quad (4.28)$$

where the J -statistic is asymptotically χ^2 distributed with $(p-l)$ degrees of freedom (over-identification).⁶ A striking feature of the method of moments approach is its transparency. In particular, it is easy to check the goodness-of-fit of the model from the moment conditions of interest, i.e. the dynamic properties of the model can be tested by evaluating graphically the match of the estimated and model-generated moments.

⁶However, if the off-diagonal components in the estimated Newey and West matrix are discarded, the the distribution in the J -statistic is likely to have a larger dispersion than the χ^2 -distribution with degrees of freedom of $p-l$. Indeed, when the weight matrix is not optimal or some moment conditions are not valid, the J -statistic is no longer χ^2 distributed. We check the validity of the weight matrix with our chosen moment conditions via a Monte Carlo study.

4.4 Empirical application to the Euro area

In this section, we first present the data for our empirical application. Then we discuss our empirical results of the structural and behavioral parameters. Finally, we examine the finite sample properties of the moment-based estimator via a Monte Carlo study and investigate three-dimensional parameter space of the BR model.

4.4.1 Data

The data source for the New Keynesian model is the 10th update of the Area-wide Model quarterly database described in Fagan et al. (2001). The output gap and interest rate gap are computed from real GDP and nominal short-term interest rate respectively using the Hodrick-Prescott filter with a standard smoothing parameter of 1600. The inflation measure is the quarterly log-difference of the Harmonized Index of Consumer Prices (HICP) instead of the GDP deflator. The inflation gap is also computed using the Hodrick-Prescott filter.⁷ The sample for this data set is available from 1970:Q1 onwards. As we use the data over five years in a rolling window analysis to estimate the perceived volatility of the output gap $\sigma(y_t)$, the data applied in this study cover the period from 1975:Q1 to 2009:Q4.

4.4.2 Basic results

We first estimate the RE and BR model parameters using the moment-based estimation presented in the previous section. Afterwards we make a comparison between the two models and examine the effects of divergence in beliefs on the inflation and output gap dynamics. As it is common in a persuasive amount of empirical studies, the discount parameter ν is calibrated to 0.99. We also fix γ to unity, which is in line with De Grauwe (2011, p. 439) and accounts for a moderate degree in the intensity of choice.⁸ By fixing those parameters in the final estimation, we can reduce problems in high-dimensional parameter space and cope with the uncertainty of the estimates. Given these assumptions, we can separately obtain the estimates for remaining parameters from the rational and bounded rationality model via the moment-based estimation. They are presented in Table 1.

⁷We resort to the HICP instead of the conceptually more appropriate implicit GDP-deflator which is common in the literature, since the former is more in line with micro data evidence. For instance, Forsells and Kenny (2004) show that inflation expectations can be approximated by micro-level data like consumer surveys (i.e. in the European Commission survey indicators). Also see Ahrens and Sacht (2011, pp. 10–11) for a more detailed discussion on using the HICP instead of the GDP-deflator in macroeconomic studies.

⁸Goldbaum and Mizrach (2008) estimated the intensity of choice parameter in the dynamic model for mutual fund allocation decision. In our application, the system with many parameters is likely to have a likelihood with multiple peaks, some of which are located in uninteresting or implausible regions of the parameter space. By fixing the intensity of choice parameter, it makes it easier to concentrate on our objective of empirical application, i.e. the interpretation of the role of bounded rationality in the NKM.

Table 4.1: Estimates of the RE and BR model

Label	RE	BR
α	0.765 (0.481 - 1.000)	0.203 (0.000 - 0.912)
χ	1.000 -	0.950 (0.000 - 1.000)
τ	0.079 (0.000 - 0.222)	0.387 (0.000 - 0.927)
κ	0.035 (0.011 - 0.058)	0.219 (0.075 - 0.362)
ϕ_y	0.497 (0.058 - 0.936)	0.673 (0.404 - 0.942)
$\phi_{\hat{\pi}}$	1.288 (1.000 - 1.944)	1.073 (1.000 - 1.775)
$\phi_{\hat{r}}$	0.604 (0.411 - 0.797)	0.673 (0.523 - 0.824)
σ_y	0.561 (0.354 - 0.768)	0.827 (0.463 - 1.190)
$\sigma_{\hat{\pi}}$	0.275 (0.097 - 0.453)	0.743 (0.449 - 1.046)
$\sigma_{\hat{r}}$	0.421 (0.140 - 0.701)	0.244 (0.000 - 0.624)
β	-	2.221 (0.000 - 9.747)
δ	-	0.665 (0.000 - 7.877)
ρ	-	0.003 (0.000 - 1.000)
J	56.30	40.30
p -value	0.8436	0.9931
5% crit. of χ^2 dist.	88.25	84.82

Note: The data cover the period spanning 1975:Q1 - 2009:Q4 (T=140 observations). The parameters ν and γ are set to 0.99 and unity, respectively. We use the rolling window of 5 years (20 observations) to compute the perceived volatility of the output gap, i.e. the unconditional standard deviation of y_t is denoted by $\sigma(y_t)$. The 95% asymptotic confidence intervals are given in brackets.

102 4. Identification of Animal Spirits in a Bounded Rationality Model

Several observations are worth mentioning. The parameter estimate of the degree of price indexation α is much higher in the RE (0.765) than the BR (0.203) model. It follows that the expressions, which are in front of the forward- and backward-looking terms in the Phillips Curve, indicate a higher weight on future inflation $\tilde{E}_t^j \hat{\pi}_{t+1}$ (i.e. $\frac{\nu}{1+\alpha\nu} > \frac{\alpha}{1+\alpha\nu}$); the result is more pronounced for the BR (0.82 > 0.18) compared to the RE model (0.56 > 0.43). For the latter, this indicates that there is strong evidence for a hybrid structure of the NKPC. The empirical applications of the BR model show that the dynamics of the inflation gap are primarily driven by the expectations (i.e. the evaluation of the forecast performance) for the inflation gap if cognitive limitation of agents is assumed. This is not necessarily true under rational expectations. In other words, we find strong evidence for an autoregressive expectation formation process, since the estimated value for α is high; one group assumes a central bank inflation target of zero percent (equation (13)), while the other group of the agents form their expectations in a purely static way (equation (14)). Regarding the dynamic IS equation, the output gap is influenced by the forward- and backward-looking terms at the same proportion, since the empirical estimates show that $\chi = 1$ and $\chi = 0.950$ hold for the RE and the BR models, respectively. In particular, this degree of habit persistence suggests that past observations strongly matter for the dynamics of the output gap. Finally, the parameter estimate for the degree of interest rate smoothing shows that there is a moderate degree of persistence ($\phi_{\hat{r},t}$) in the nominal interest rate gap for both models.

Furthermore, while the empirical estimates for κ and τ in the RE model indicate a small degree of inherited persistence due to changes in the real interest rate gap and the output gap respectively, this does not hold for the BR model. Here the changes in the output gap have a strong impact ($\kappa = 0.219$) on movements in the inflation gap relative to the RE case ($\kappa = 0.035$). For the output gap, inherited persistence plays a fundamental role in shaping the dynamics of this economic indicator, which can be seen through the high values of inverse intertemporal elasticity of substitution. For the BR model, this value ($\tau = 0.387$) is much larger than the one for the RE model ($\tau = 0.079$). This implies that the tendency towards risk aversion in the BR is stronger than the RE model. To sum up, our results show that in the BR model cross-movements in the output and inflation gap account for persistence in both variables (under consideration of perfect habit formation $\chi = 1$) rather than price indexation alone. This can be seen through the high values of κ and τ compared to α . For the RE model, the opposite holds.

The output and inflation gap shocks, whose magnitudes are estimated to be $\sigma_y = 0.827$ and $\sigma_{\hat{\pi}} = 0.743$ respectively, are larger for the BR than those of the RE model. The results reveal that the volatilities of the output and inflation gap are strengthened by the effects of behavioral heterogeneity on the consumption and pricessetting rules. For instance, the waves of optimism and pessimism act as a persistent force in the output gap fluctuations with peaks and troughs. Figure 1 illustrates that the peak of the fluctuation in the simulated output gap (middle-left panel) corresponds to the market optimism (lower-left panel) and vice versa. The qualitative interpretation remains almost the same for the inflation gap dynamics

(middle- and lower-right panel respectively) - but the dynamics of extrapolators are highly volatile reflecting the large second moment of the empirical inflation gap (upper-right panel). The goodness-of-fit of the models could not be directly compared by illustrating the simulated time series (middle-panels), but we can see that the series resemble qualitatively their empirical counterparts (upper-panels). Finally, the nominal interest rate shocks $\sigma_{\hat{r}}$ in the RE model are estimated to be roughly twice as large as in the BR model.

The remaining parameter estimates confirm the known results from the literature where the monetary policy coefficient on the output gap is low while the opposite holds for the coefficient on the inflation gap. The latter indicates that the Taylor principle holds over the whole sample period. Nevertheless, the results for the BR model indicate a stronger concern in the output gap movements relative to the dynamics in the inflation gap. Again, the opposite is true for the RE model. It is worth mentioning that the estimation results indicate a monetary policy coefficient on the output gap ϕ_y of 0.673, which is in line with the observations of De Grauwe (2011, pp. 443-445). His simulations show that flexible inflation targeting can reduce both output gap and inflation (gap) variability at a minimum level if ϕ_y lies in the range of 0.6 to 0.8.

The interpretation of this observation is two-fold. First, consider the case of strict inflation targeting, where the central bank does not account for the volatility in the output gap. As a result, the forecast performance of the optimists and pessimists are not affected since the (real) interest rate gap in the dynamic IS curve does not response directly to monetary policy. However, there is still an indirect effect (even highly volatile movements in y_t are not dampened by the policy makers) indicated by κ in the NKPC. Hence, due to the high degree of inherited persistence the strict inflation targeting can fail to control strong fluctuations in the output and inflation gap. Second, in the case of strong output gap stabilization (relative to the inflation gap) the central bank dampens its pre-commitment to an inflation target. The amplification effects of this kind of policy on the forecast performances of the inflation extrapolators will then result in higher inflation variability. We conclude that our empirical findings account for neither the first nor the second extreme case, but for a optimal flexible inflation targeting in the Euro Area over the observed time interval instead.

As already noted, the present study focuses on the estimation of the bounded rationality parameters. First, we come to the conclusion that over the whole sample period, the optimistic agents have expected a fixed divergence of belief of $\beta = 2.221$. Roughly speaking, the optimists have been really optimistic that the future output gap will differ *positively* by slightly above one percent on average from its steady state value.⁹ Due to the symmetric structure of the divergence in beliefs, over the same sample period pessimistic agents were moderately pessimistic instead, since from their point of view the future output gap was expected to be around one

⁹Note that expected future value of the output gap is given by $E_t^i y_{t+1} = |d_t| = \frac{1}{2}\beta$ on average with $i = \{O, P\}$.

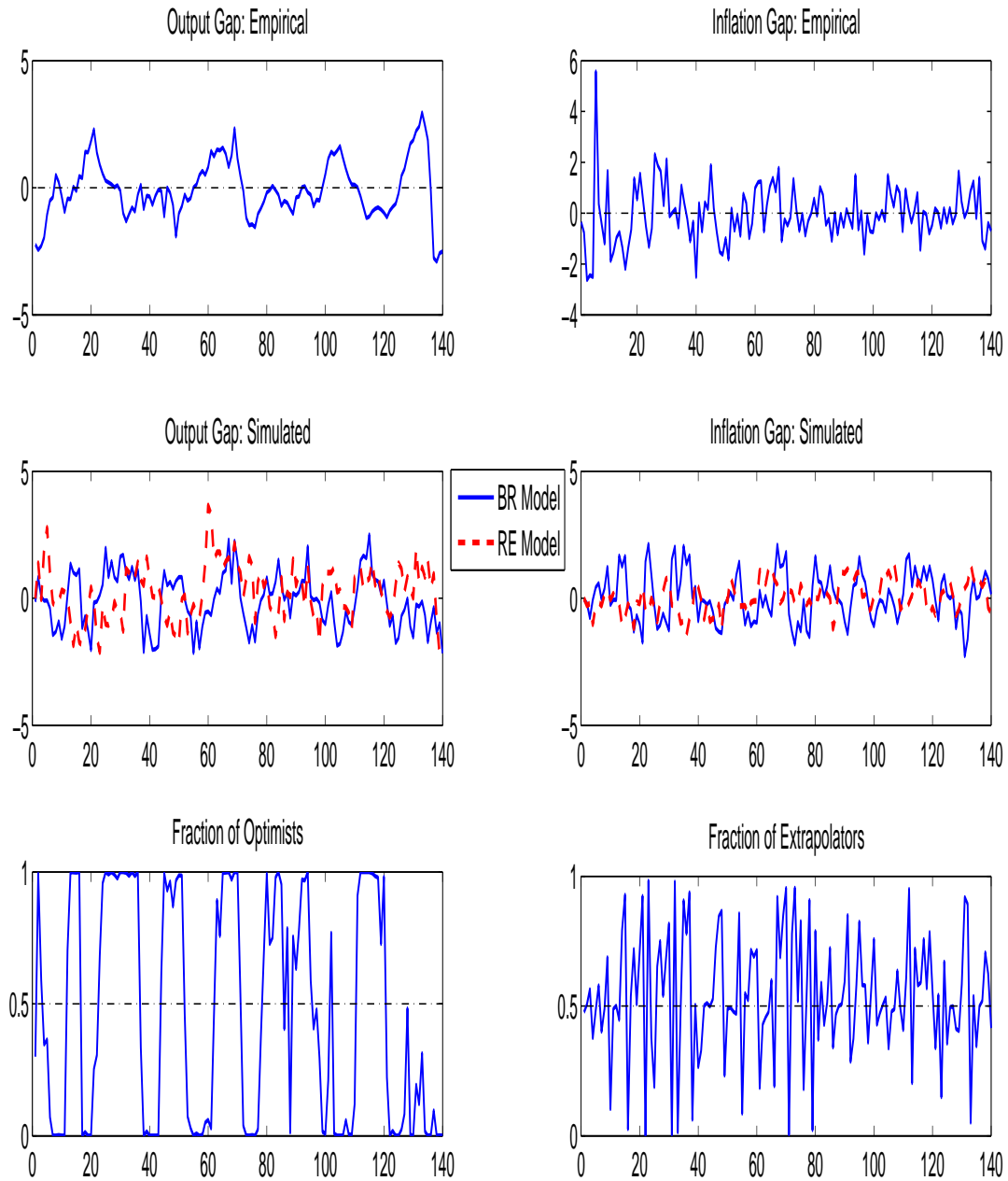


Figure 4.1: Dynamics in the output gap and the inflation gap.

Note: Upper and middle panels plot empirical and simulated values for the output gap (left) and the inflation gap (right), while lower panels plot the corresponding fraction of market optimists (left) and extrapolators (right). The simulated time series are computed using the parameter estimates for both models given in Table 1.

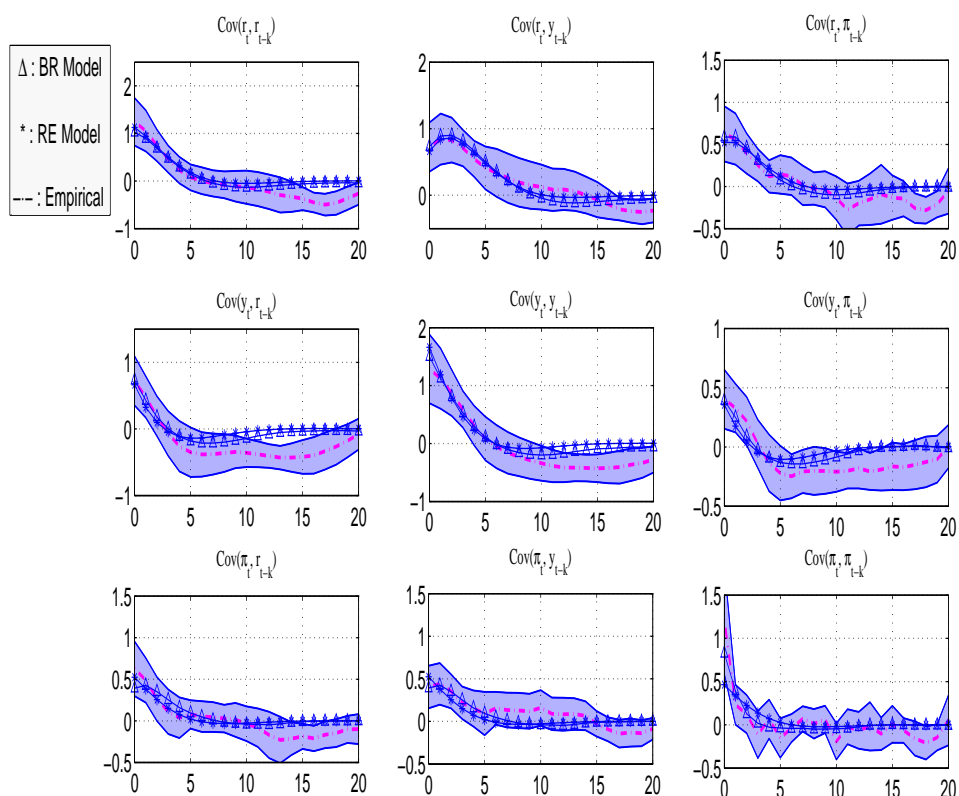


Figure 4.2: Model covariance (Cov) profiles in the Euro Area.

Note: The dashed line results from the empirical covariance estimates. The shaded area is the 95% confidence bands around the empirical moments. The triangle (BR) and star (RE) lines indicate the model generated ones. The confidence bands are computed via the Delta method (see Appendix B).

percent on average *below* its steady state value. Furthermore, both types of agents felt safely about their expectations due to the fact that the estimate for the variable component in the divergence of pessimistic beliefs is very low ($\delta = 0.665$) - this implies that there is a low degree of uncertainty connected to the expected future value of y_t . In line with the results for (and assumptions of) the parameters, which indicate endogenous and inherited persistence (α, χ, κ and τ), the highly subjective expected mean value of the output gap β - in conjunction with the dynamics induced by the self-selecting mechanisms (see the corresponding fractions in the lower-panels in Figure 1) - explains the high volatility of the output gap. Based on discrete choice theory, this strengthens the optimistic agents' belief about the future output gap to diverge in the data, since they can over(or under)-react to underlying shocks that occur across the Euro Area. The same observation holds for the inflation gap dynamics. The proportion of the extrapolators in the economy corresponds to the inflation gap movements (cf. lower right vs. upper-right panels in Figure 1): the higher the fraction of extrapolators is, the more volatile the inflation gap dynamics will be. Finally, ρ is estimated to be zero, i.e. past errors are not taken into account (cf. equations (4.10) and (4.16)). This leads to the conclusion that strict forgetfulness or cognitive limitation holds, which is a requirement for observing animal spirits (cf. De Grauwe (2011, p. 440)).

Indeed, visual inspection shows a fairly remarkable goodness-of-fit of the models to data (see Figure 2). The match both models achieve looks clearly good over the first few lags and still fairly good over the higher lags until the lag 8. In any case, all of the moments are now inside the confidence intervals of the empirical moments. This even holds true for some covariances up to lag 20. This is also confirmed by the values of the loss function J for the RE (56.30) and BR (40.30) model given in the last row of Table 1. The asymptotic χ^2 distributions for the J -test have the degrees of freedom of 68 and 65 for the RE and BR model, respectively. Since the critical values at 5% level are 85.25 and 84.82 respectively, and the estimated loss function values are smaller than these criteria, we do not reject the null hypothesis that these models are valid. Moreover, the picture shows a remarkable fit of the BR model, which leads to some confidence in the estimation procedure. We conclude that a bounded rationality model with cognitive limitation provides good fits for auto- and cross-covariances of the data.

Note here that the significant differences between two models have to be tested by a formal model comparison method, since the models do not have any difficulties to fit the empirical moments at the 5% significant interval (see also Jang (2012) among others). In other words, the J -test only evaluate the validity of the model along the lines of the chosen moment conditions. Therefore we cannot provide a direct comparison between the fits of the two models. More rigorous test will be a priority for future research.

Finally, our empirical results indicate that the empirical test of bounded rationality (viz. the assumption of the divergence in beliefs) has to be treated carefully, because all parameters (especially the behavioral ones) within the non-linear modeling approach are generally poorly determined, i.e. wide confidence bands occur.

We delve into this problem by examining the finite size properties of the moment-based procedure through a Monte Carlo study and a sensitivity analysis presented in the next section. Our results from these exercises will achieve confidence in the parameter estimates given in Table 1.

4.4.3 Comparison with other studies

There exists a plethora of studies on the estimation of (small, medium or large) NKM with rational expectations using Euro Area data. However, to the best of our knowledge these studies are different to our contribution in several dimensions. While we apply a moment-based estimation on the Euro Area data over a specific time interval up to the end of 2009, most of the investigations are based on the generalized method of moments and Bayesian estimations using data just to the beginning of the 21st century instead. Furthermore, we consider gap specifications of $\hat{\pi}_t$ and \hat{r}_t explicitly while in the literature the majority of time series are not detrended. Hence, a comparison of our results with those from the literature has to be done with some caution.

More generally, one of the representative studies in this field is the empirical application of Smets and Wouters (2003). Here the sample period captures the period from 1980:Q2 to 1999:Q4. In their chapter, they apply Bayesian estimation on a medium scale model for the Euro Area. Compared to the cases of the RE and BR presented here, they found different values for the parameters τ and $\phi_{\hat{\pi}_t}$, which are estimated to be higher (0.739 and 1.684). In contrast, the estimated values for κ and ϕ_y are relatively small (0.01 and 0.10). Finally, $\phi_{\hat{r}} = 0.673$ is slightly lower than in Smets and Wouters (2003, $\phi_r = 0.956$).

Moons et al. (2007) give a good overview on the results stemming from different studies using different techniques except for the Bayesian one. Most of the parameter estimates are in line with those reported in column 1 of our Table 1, i.e. in case of the RE model. According to Table 1 in Moons et al. (2007, p. 888) τ and κ vary in a range of (0.03, 0.08) and (0.02, 0.17), while we find $\tau = 0.079$ and $\kappa = 0.035$. The results for the policy parameters $\phi_{\hat{y}} = 0.604$, $\phi_{\hat{\pi}} = 1.288$ and $\phi_{\hat{r}} = 0.497$ are slightly below the estimates reported in Moons et al. (2007) where $\phi_y = (0.77, 0.90)$, $\phi_{\pi} = (0.87, 2.02)$ and $\phi_r = (1, 3.2)$. For the latter, note once again that the level and not the gap of the corresponding time series is considered. The composite parameter, which indicate backward-looking behavior in the dynamic IS curve and the NKPC, can be denoted by $\psi_1 = \frac{\chi}{1+\chi}$ and $\psi_2 = \frac{\alpha}{1+\alpha\nu}$. It can be stated that our results for the RE model, $\psi_1 = 0.5$ and $\psi_2 = 0.43$, mimic roughly those found in the literature, i.e. $\psi_1 = (0.22, 0.97)$ and $\psi_2 = (0.13, 0.54)$.

Comparing the results discussed in the previous paragraph with those presented in column 2 of Table 1, it can be seen that in the case of the BR model these results differ substantially from the those reported in the literature. Not surprisingly, this stems from the fact that the behavioral model of De Grauwe exhibits a different kind of expectation channel which can substitute the absence of rational expectations for the model dynamics. Nevertheless, Moons et al. (2007) estimate a small scale

NKM of an open-economy under consideration of a fiscal policy rule (in the spirit of the European Stability and Growth Pact) with Bayesian techniques and found the parameter estimates, which are similar with our results. In particular, τ is estimated to be high (0.24) which is in line with the BR model (0.387). The authors also find that a high value of the monetary policy coefficient concerning the output gap is estimated to be $\phi_y = 0.75$, while we find a value of 0.673.

4.5 Robustness checks

In this section, we report the variation of the parameter estimates under both the RE and BR model. First, we study the finite size properties of the moment-based estimation using the Monte Carlo study. The result shows that we can reduce the estimation uncertainty presented here with a large sample size. Compared to the RE model, however, the parameter estimates of the BR model have wide confidence intervals, because the non-linearity of the model gives rise to additional parameter uncertainty during the estimation. This affects the corresponding values of the bounded rationality parameters β , δ and the memory parameter ρ in the forecasting heuristics (4.11) and (4.12) as well as (4.17) and (4.18). Second, we investigate the sensitivity of these behavioral parameters in the objective function by investigating three-dimensional parameter space. We vary these parameters in a reasonable range to find the lowest value of the loss function (4.21).

4.5.1 Monte Carlo study

To analyze the finite sample properties in the macro data, we use three sampling periods in the data generating process ($T=100, 200, 500$). The experimental true parameters are drawn from the parameter estimates in the previous section. After 550 observations are simulated, we discard the first 50 observations to trim a transient period. In the RE model, we compute the empirical moment conditions and its Newey-West weight matrix of each artificial time series, and estimate the parameters using the method of moment estimator over 500 replications. The same procedure is carried out to estimate the parameters of the BR model. However, this makes the computation expensive for the simulated method of moment estimator. We reduce the computational cost by setting the simulation size to $S = 10$.¹⁰

Table 4.2 summarizes the results from the MC experiment for the RE model. We report the mean, the root mean square error (RMSE) and the standard error (S.E). The true values of the parameters are stated in the second column. The results show that the method of moment estimation of the RE model has good finite sample properties; see the RMSE sensitivity to variations in sample size. The large sample size remarkably improves the asymptotic efficiency of the method of

¹⁰The implementation of the MC study on the model with a large simulation size (i.e. $S=100$) does not have a drastic change in parameter estimates; see appendix C. The approximation error rates of analytic moments are 10% and 1% for the simulation sizes $S = 10$ and 100, respectively. The computation becomes expensive when the large simulation size is used.

Table 4.2: Monte Carlo study for the RE model

Label	True (θ^0)	T=100		T=200		T=500	
		Mean	RMSE	Mean	RMSE	Mean	RMSE
α	0.750	0.802	0.174	0.778	0.125	0.763	0.079
		S.E: 0.155		S.E: 0.112		S.E: 0.073	
χ	1.000	0.943	0.128	0.939	0.127	0.946	0.103
		S.E: 0.365		S.E: 0.293		S.E: 0.202	
τ	0.085	0.100	0.062	0.088	0.043	0.083	0.029
		S.E: 0.079		S.E: 0.061		S.E: 0.041	
κ	0.035	0.047	0.026	0.042	0.016	0.039	0.009
		S.E: 0.016		S.E: 0.011		S.E: 0.071	
ϕ_y	0.500	0.518	0.267	0.487	0.167	0.487	0.107
		S.E: 0.236		S.E: 0.162		S.E: 0.104	
$\phi_{\hat{\pi}}$	1.250	1.350	0.309	1.322	0.217	1.296	0.146
		S.E: 0.343		S.E: 0.222		S.E: 0.144	
$\phi_{\hat{r}}$	0.600	0.623	0.111	0.615	0.076	0.611	0.046
		S.E: 0.094		S.E: 0.069		S.E: 0.045	
σ_y	0.600	0.632	0.127	0.627	0.090	0.623	0.059
		S.E: 0.125		S.E: 0.095		S.E: 0.062	
$\sigma_{\hat{\pi}}$	0.275	0.249	0.075	0.263	0.049	0.270	0.030
		S.E: 0.062		S.E: 0.046		S.E: 0.031	
$\sigma_{\hat{r}}$	0.400	0.234	0.240	0.289	0.181	0.345	0.105
		S.E: 7.487		S.E: 1.456		S.E: 1.026	
J		30.58		24.12		20.10	
# of rejections		4		6		0	

Note: ν is set to the value of 0.99. The reported statistics are based on 500 replications. RMSE is the root mean square error. S.E denotes the mean of standard error.

moments estimator, since the mean of standard error for the estimates becomes the smallest. However, the estimated value for the policy shock parameter σ_r often hit the boundary (i.e. $\sigma_r = 0.0$) and makes the numerical derivative of the moment conditions unstable. This leads to the large asymptotic error for this parameter.¹¹

The J -statistic is used to evaluate the validity of the two models when fitting the artificial data. On average, the null hypothesis that the model is the true one is not rejected according to the over-identification test for both the RE and BR model; e.g. for the sample size of $T=100$, the J test rejects the validity of the RE and BR model for 4 and 16 times, respectively. The number of rejection is very small, since the simulated replications are 500. And we do not find any rejection of both models when a large sample size is used ($T=500$). In addition, it can be seen from the J test that the BR model fits the data slightly better than the RE model on average. Nevertheless, the direct diagnostic comparison between two models must be made with caution, because the BR model has more parameters than the RE model, i.e. their χ^2 -distributions are different.

In comparison with the results of the RE model, we found that the simulated method of moments regarding the BR model has more or less poor finite sample properties when inspecting the parameters α , τ , β , and δ (see Table 4.3). However, the large uncertainty for the parameter estimates can be mitigated by more observations in the data. On the other side, note here that we can consistently recover the true values for the other parameter estimates. Put differently, the parameter estimates almost converge to the true ones as the sample size increases (i.e. $T=500$). In this case the RMSE gets smaller. The large sample allows us to make more accurate inference about the group behavior in the market expectation formation processes. Indeed, as market behavior is unobservable in most cases, we need a large sample size to consistently estimate the behavioral parameters. Nevertheless, the estimated results for the behavioral parameters can be seen as confident starting values used for calibration exercises like for e.g. (optimal) monetary and fiscal policy analysis.

4.5.2 Sensitivity of the behavioral parameters

In this sensitivity analysis we investigate the region of the objective function with respect to different values of β , δ and ρ . The findings from the MC study indicate that the RMSE values for these behavioral parameters in the discrete choice theory are higher than those for the other structural parameters even for a large sample size. We discuss the poor finite sample properties of these crucial parameters in the BR model by evaluating the loss function under consideration of different pairs for β , δ and ρ . The remaining parameters are fixed on their estimated values taken from the second column of Table 1. It is our aim to pin down those values from the

¹¹Note here that we use the optimization tool (Matlab version R2010a) with the *fmincon* solver. Especially the interior-point algorithm has a number of advantages over other algorithms (i.e., active-set, trust-region-reflective, and sqp). For example, the implementation of the interior-point algorithm for large-scale linear programming is considerably simpler than for the other algorithms. Furthermore, it can handle nonlinear non-convex optimization problems of the BR model.

Table 4.3: Monte Carlo study for the BR model

Label	True (θ^0)	T=100		T=200		T=500	
		Mean	RMSE	Mean	RMSE	Mean	RMSE
α	0.200	0.326	0.286	0.383	0.278	0.285	0.187
		S.E: 0.312		S.E: 0.232		S.E: 0.142	
χ	1.000	0.666	0.724	0.798	0.679	0.850	0.655
		S.E: 1.802		S.E: 1.614		S.E: 1.470	
τ	0.385	1.075	0.837	0.620	0.370	0.550	0.292
		S.E: 0.810		S.E: 0.341		S.E: 0.216	
κ	0.215	0.246	0.153	0.223	0.142	0.225	0.139
		S.E: 0.076		S.E: 0.051		S.E: 0.035	
ϕ_y	0.675	0.757	0.458	0.697	0.435	0.694	0.430
		S.E: 0.248		S.E: 0.116		S.E: 0.065	
ϕ_{π}	1.100	1.090	0.703	1.069	0.699	1.089	0.699
		S.E: 0.326		S.E: 0.174		S.E: 0.106	
$\phi_{\hat{r}}$	0.670	0.681	0.427	0.675	0.425	0.681	0.424
		S.E: 0.073		S.E: 0.046		S.E: 0.028	
σ_y	0.825	0.872	0.549	0.888	0.533	0.874	0.527
		S.E: 0.290		S.E: 0.182		S.E: 0.133	
σ_{π}	0.740	0.606	0.496	0.647	0.477	0.699	0.470
		S.E: 0.090		S.E: 0.053		S.E: 0.034	
$\sigma_{\hat{r}}$	0.240	0.165	0.182	0.180	0.176	0.165	0.180
		S.E: 0.169		S.E: 0.140		S.E: 0.113	
β	2.250	2.831	1.867	2.440	1.608	2.330	1.543
		S.E: 5.638		S.E: 4.149		S.E: 3.670	
δ	0.650	1.293	1.021	0.925	0.750	0.862	0.663
		S.E: 4.000		S.E: 3.596		S.E: 3.223	
ρ	0.000	0.213	0.218	0.104	0.134	0.093	0.122
		S.E: 0.422		S.E: 0.386		S.E: 0.335	
J		29.34		22.30		20.74	
# of rejections		16		1		0	

Note: ν is set to the value of 0.99. The reported statistics are based on 500 replications. RMSE is the root mean square error. S.E denotes the mean of standard error.

112 4. Identification of Animal Spirits in a Bounded Rationality Model

parameter space, which are associated with the lowest value of the loss function.

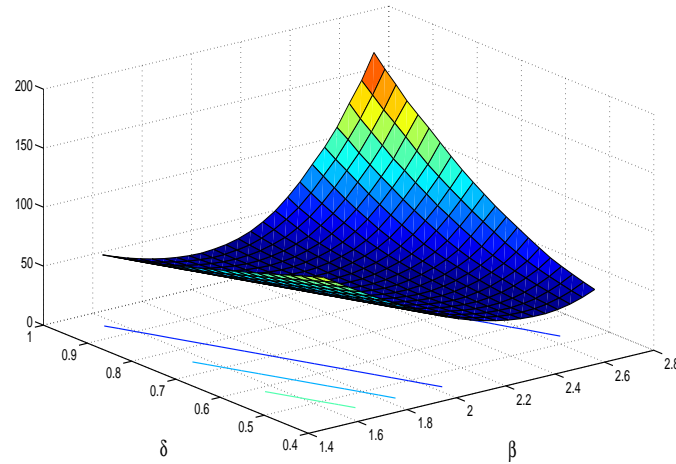


Figure 4.3: 3-D contour plot of the parameter space with β and δ

Note: The value of the quadratic objective function J is given on the vertical axis.

Figures 3 to 5 illustrate three contour plots, from which we can examine the region of the loss function J under consideration of the pairwise variation in all three parameters over a reasonable range. We see from Figure 3 that the minimum value of the loss function is centered around $(\delta, \beta) = (0.6, 2.2)$. This observation is in line with our results given in Table 1, and indicates that applying the method of moment approach leads to consistent parameter estimates. However, our result emphasizes that the shape of the contour plot is moderately flat for specific combinations of δ and β , i.e. which still indicates the existence of wide confidence bands. Note that the value of the loss function increases dramatically if δ and β deviate strongly from their estimated values. In this case a trade-off arises: a highly predicted subjective mean value β requires a low degree of divergence δ in order to ensure a minimum value of J .

Figure 4 and 5 show that the minimum of the loss function is given by a value of the memory parameter ρ equal to zero in conjuncture with the estimated values of β and δ around 2.2 and 0.6, respectively. This result confirms the estimate of ρ given in Table 1 and strengthens our argumentation in section 4 since strict forgetfulness holds as a requirement for observing animal spirits.

In sum, this simulation results show that for a small sample size, the results from a MC study and a sensitive analysis confirm the absence of statistical accuracy of these behavioral parameters (i.e. the case of wide confidence bands) when applying the method of moment approach.

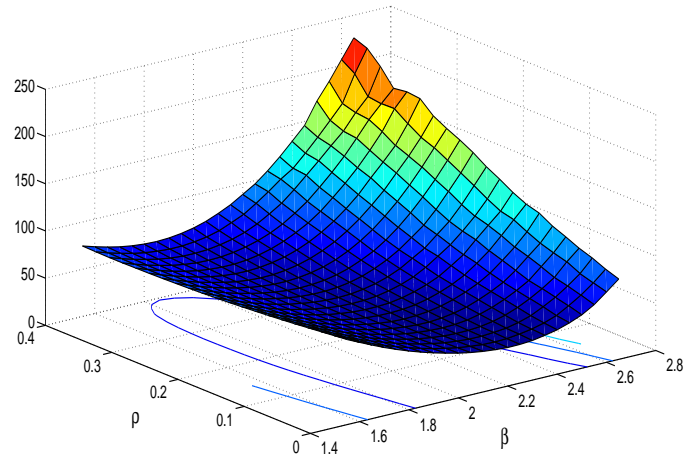


Figure 4.4: 3-D contour plot of the parameter space with β and ρ

Note: The value of the quadratic objective function J is given on the vertical axis.

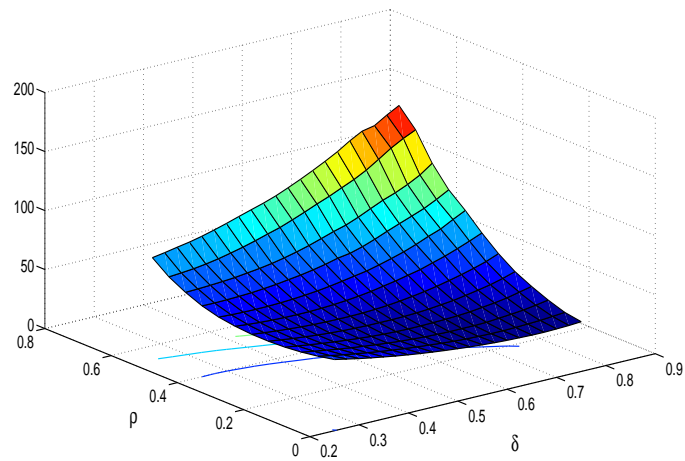


Figure 4.5: 3-D contour plot of the parameter space with δ and ρ

Note: The value of the quadratic objective function J is given on the vertical axis.

4.6 Conclusion

In this chapter, we attempt to provide empirical evidence for the behavioral assumption in the model of De Grauwe (2011). The validity of the model assumption on the cognitive limitation (e.g. because of different individual emotional states) is empirically tested using historical Euro Area data. We attempt to identify the so-called behavioral parameters, which account for animal spirits in the Euro Area; i.e. we hypothesize that historical movements of macro dynamics are influenced by waves of optimism and pessimism.

To examine the effects of the group behavior on the output and inflation gap, we follow the behavioral approach of De Grauwe (2011), who assumes divergence in beliefs about the future value of both variables. The corresponding decision rules for market optimism and pessimism are given by the forecast performance of the agents from the discrete choice theory. To see this, we contrast a standard hybrid version of the three-equations New-Keynesian model of rational expectations with a version of the same model where we assume bounded rationality in expectation formation processes using the moment-based estimation.

Our main empirical findings show that a bounded rationality model with cognitive limitation provides a reasonable fit to auto- and cross-covariances of the Euro Area data. Therefore our empirical results of the BR model offer some new insights into expectation formation processes for the Euro Area. First, over the whole time interval the agents had expected moderate deviations of the output gap from its steady state value with low uncertainty. Second, in the absence of rational behavior we find strong evidence for an autoregressive expectation formation process regarding the inflation gap. Both observations explain a high degree of persistence in the output gap and the inflation gap. Based on the discrete choice theory and the self-selection process of the agents, we found that animal spirits strengthen the optimistic's belief about the future output gap to diverge in the historical Euro Area data.

To the best of our knowledge, such kind of empirical studies have not been extensively investigated before in the literature. However, the empirical test of bounded rationality (viz. the assumption of the divergence in beliefs) has to be treated carefully, because the parameters (especially the behavioral ones) within the non-linear modeling approach are poorly determined, i.e. wide confidence bands occur. We delve into this problem by examining the finite size properties of the moment-based procedure through a Monte Carlo study and a sensitivity analysis. In the end, we provide empirical evidence in support of De Grauwe (2011, fn. 4) for understanding the group's over- and under-reaction to the economy. In order to identify the effects of individual expectation formation processes on the economy, in further research, the decision rules i.e. the transition rules from one state of the economy to another can be calculated based on survey data (for example see Lux (2009)). Thus these probabilities are then treated as exogenous and (in contrast to the De Grauwe model) are computed under consideration of the underlying time series using the discrete choice theory. Finally and only if the estimation of small-

scale models is considered to be satisfactory, one can further continue the model estimation with much richer models like e.g. the medium-scale version developed by the Smets and Wouters (2005, 2007). We leave these issues to future research.

Appendix

A: Solution of the NKM

In general, all model specifications are described by the following system in canonical form:

$$AX_t + BX_{t-1} + CX_{t+1} + \varepsilon_t = 0 \quad (4.29)$$

where

$$X_t = \begin{pmatrix} y_t \\ \hat{\pi}_t \\ \hat{r}_t \end{pmatrix}, X_{t-1} = \begin{pmatrix} y_{t-1} \\ \hat{\pi}_{t-1} \\ \hat{r}_{t-1} \end{pmatrix}, X_{t+1} = \begin{pmatrix} \tilde{E}_t^j y_{t+1} \\ \tilde{E}_t^j \hat{\pi}_{t+1} \\ \tilde{E}_t^j \hat{r}_{t+1} \end{pmatrix}, \varepsilon_t = \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{\hat{\pi},t} \\ \varepsilon_{\hat{r},t} \end{pmatrix}.$$

The corresponding system matrices are given by:

$$A = \begin{pmatrix} 1 & 0 & \tau \\ -\lambda & 1 & 0 \\ -\phi_{\hat{r}}\phi_y & -\phi_{\hat{r}}\phi_{\pi} & 1 \end{pmatrix}, B = \begin{pmatrix} -\frac{\chi}{1+\chi} & 0 & 0 \\ 0 & -\frac{\alpha}{1+\alpha\nu} & 0 \\ 0 & 0 & -(1-\phi_{\hat{r}}) \end{pmatrix} \quad (4.30)$$

and

$$C = \begin{pmatrix} -\frac{1}{1+\chi} & -\tau & 0 \\ 0 & -\frac{\nu}{1+\alpha\nu} & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (4.31)$$

Recall that for the rational expectations model we assume

$$\begin{aligned} \tilde{E}_t^{RE} y_{t+1} &= E_t y_{t+1} \\ \tilde{E}_t^{RE} \hat{\pi}_{t+1} &= E_t \hat{\pi}_{t+1} \end{aligned}$$

and for the bounded rationality model we assume

$$\begin{aligned} \tilde{E}_t^{BR} y_{t+1} &= (\alpha_{y,t}^O - \alpha_{y,t}^P) d_t \\ \tilde{E}_t^{BR} \hat{\pi}_{t+1} &= \alpha_{\hat{\pi},t}^{tar} \bar{\hat{\pi}} + \alpha_{\hat{\pi},t}^{ext} \hat{\pi}_{t-1} \end{aligned}$$

where we also consider equations (4.10) to (4.18) with $\bar{\hat{\pi}} = 0$. In the following, we solve for the dynamics of the system (4.29). In case of the BR model, the solution is given by

$$X_t = -A^{-1}[BX_{t-1} + CX_{t+1} + \varepsilon_t] \quad (4.32)$$

where the matrix A is of full rank, i.e. its determinant is not equal to zero, given the parameter estimates in section 4. Under consideration of the heuristics for the forecasts regarding the output and inflation gap expectations, the forward looking term

X_{t+1} is substituted by the equivalent expressions for the discrete choice mechanism given in section 2. It follows that the model becomes purely backward-looking and thus (4.32) can be solved by backward-induction.

In contrast, the RE model is both backward- and forward-looking. Therefore we apply the method of undetermined coefficients in order to solve the model. The law of motion which describes the analytical solution is given by

$$X_t = \Omega X_{t-1} + \Phi \varepsilon_t \quad (4.33)$$

where $\Omega \in \mathbb{R}^{3 \times 3}$ and $\Phi \in \mathbb{R}^{3 \times 3}$ are the solution matrices. The former is a stable matrix as long as (similar to the matrix A in the BR case) its determinant is not equal to zero, which ensures the invertibility of Ω . Again, this is confirmed given the estimation results in section 4. We substitute (4.33) into (4.29) which yields

$$A(\Omega X_{t-1} + \Phi \varepsilon_t) + B X_{t-1} + C(\Omega X_t + \Phi E_t \varepsilon_{t+1}) + \varepsilon_t = 0.$$

This is equivalent to

$$A(\Omega X_{t-1} + \Phi \varepsilon_t) + B X_{t-1} + C(\Omega^2 X_{t-1} + \Omega \Phi \varepsilon_t + \Phi E_t \varepsilon_{t+1}) + \varepsilon_t = 0.$$

Hence, the reduced form can be rewritten as

$$(C\Omega^2 + A\Omega + B)X_{t-1} + (A\Phi + C\Omega\Phi + I)\varepsilon_t = 0 \quad (4.34)$$

with I being the identity matrix. Note that $\varepsilon_t \sim N(0, \sigma_z^2)$ with $z = \{y, \hat{\pi}, \hat{r}\}$ and thus $E_t \varepsilon_{t+1} = 0$. In order to solve equation (4.34), all the terms in brackets must be zero.¹² Thus the solution matrices can be uniquely determined. We may write that as

$$C\Omega^2 + A\Omega + B = 0 \Rightarrow \Omega = -(C\Omega + A)^{-1}B. \quad (4.35)$$

In order to solve the quadratic matrix equation (4.35) numerically, we employ the brute force iteration procedure mentioned in Binder and Pesaran (1995, p. 155, fn 26). Hence an equivalent recursive relation of (4.35) is given by

$$\Omega_n = -(C\Omega_{n-1} + A)^{-1}B \quad (4.36)$$

with an arbitrary number of iteration steps N where $n = \{1, 2, \dots, N\}$. We define $\Omega_0 = \xi I$ with $0 \leq \xi \leq 1$. The iteration process (4.36) proceeds until $\|\Omega_n - \Omega_{n-1}\| < \xi$ where ξ is an arbitrarily small number. Given the solution of Ω , the computation of Φ is straightforward:

$$A\Phi + C\Omega_n\Phi + I = 0 \Rightarrow \Phi = -(A + C\Omega_n)^{-1}. \quad (4.37)$$

¹²Obviously the trivial solution $X_{t-1} = \Gamma_t = \varepsilon_t = 0$ is discarded.

B: Delta method and confidence interval for auto- and cross-covariances

The Delta method is a common technique for providing the first-order approximations to the variance of a transformed parameter; see chapter 5 of Davidson and Mackinnon (2004) among others. In the study, we use the Delta method when computing the standard errors of the estimated auto- and cross-covariances of the data. The covariance is defined as follows:

$$\gamma_{ij}(h) = E[(X_{i,t} - \mu_i)(X_{j,t+h} - \mu_j)'], \quad t = 1, \dots, T \quad (4.38)$$

where γ_{ij} is the auto-covariance function when $i = j$. Otherwise γ_{ij} denotes the cross-covariance between $X_{i,t}$ and $X_{j,t+h}$. h denotes the lag in data and μ_i (or μ_j) is the sample mean of the variable X_i (or X_j). The covariance function in Equation (4.38) proceeds with a simple multiplication:

$$\gamma_{ij}(h) = E[X_{i,t} \cdot X'_{j,t+h}] - \mu_i \cdot E[X'_{j,t+h}] = \mu_{ij} - \mu_i \cdot \mu_j$$

where μ_{ij} denotes $E[X_{i,t} \cdot X'_{j,t+h}]$. Now we see that $\gamma_{ij}(h)$ is a transformed function of the population moments μ_i , μ_j and μ_{ij} . Denote the vector μ as the collection of the moments: $\mu = [\mu_i \quad \mu_j \quad \mu_{ij}]$. We differentiate the covariance function with respect to the vector μ :

$$D = \frac{\partial \gamma_{ij}(h)}{\partial \mu} = \begin{bmatrix} \frac{\partial \gamma_{ij}(h)}{\partial \mu_i} \\ \frac{\partial \gamma_{ij}(h)}{\partial \mu_j} \\ \frac{\partial \gamma_{ij}(h)}{\partial \mu_{ij}} \end{bmatrix} = \begin{bmatrix} -\mu_j \\ -\mu_i \\ 1 \end{bmatrix} \quad (4.39)$$

Therefore the Delta method provides the asymptotic distribution of the estimate $\hat{\gamma}_{ij}$ by matching the sample moments of the data.

$$\sqrt{T}(\gamma_{ij} - \hat{\gamma}_{ij}) \sim N(0, D'SD). \quad (4.40)$$

For some suitable lag length q , we use a common HAC estimator of Newey and West (1994) when estimating the covariance matrix of sample moments. Specifically, we follow the advice in Davidson and MacKinnon (2004, p.364) and scale q with $T^{1/3}$. Accordingly we may set $q = 5$ for the Euro area data.

$$\hat{\Sigma}_\mu = \hat{C}(0) + \sum_{k=1}^q \left(1 - \frac{k}{q+1}\right) [\hat{C}(k) + \hat{C}(k)'] \quad (4.41)$$

$$\hat{C}(k) = \frac{1}{T} \sum_{t=k+1}^T [f(z_t) - \hat{\mu}][f(z_{t-h}) - \hat{\mu}]'$$

where $f(z_t) = [X_i, X_j, X_i \cdot X_j]$. We use the optimal weight matrix $S = \hat{\Sigma}_\mu^{-1}$ in estimating the covariance matrix of moments. Let s_γ be $\sqrt{D'SD}$. Then the 95% asymptotic confidence intervals for auto- and cross-covariance estimates become:

$$[\gamma_{ij} - 1.96 \cdot s_\gamma, \quad \gamma_{ij} + 1.96 \cdot s_\gamma]. \quad (4.42)$$

C: Large-scale simulation study for the BR model

We report the results of a simulation study for the BR model when a large simulation size is used; $S=100$. At present, we see that the model estimates using a large simulation size have slightly smaller values for the RMSEs than ones from a small simulation size in the section 4.3.

Table 4.4: Monte Carlo Study for the BR Model

T=100			
Label	True (θ^0)	Mean	RMSE
α	0.200	0.249	0.262 S.E: 0.314
χ	1.000	0.693	0.716 S.E: 2.653
τ	0.385	0.884	0.818 S.E: 0.699
κ	0.215	0.236	0.154 S.E: 0.0800
ϕ_y	0.675	0.728	0.454 S.E: 0.181
$\phi_{\hat{\pi}}$	1.100	1.105	0.701 S.E: 0.701
$\phi_{\hat{\tau}}$	0.670	0.677	0.427 S.E: 0.066
σ_y	0.825	0.913	0.561 S.E: 0.302
$\sigma_{\hat{\pi}}$	0.740	0.689	0.479 S.E: 0.081
$\sigma_{\hat{\tau}}$	0.240	0.165	0.198 S.E: 0.246
β	2.250	2.585	1.812 S.E: 9.275
δ	0.650	1.139	0.992 S.E: 7.656
ρ	0.000	0.224	0.239 S.E: 0.547
J		28.53	
# of rejections		9	

Note: ν is set to the value of 0.99. The reported statistics are based on 500 replications. RMSE is the root mean square error. S.E denotes the mean of standard error. Since the simulation studies become computationally expensive with a large sample size, we only report the case of $T = 100$.

PART II. Financial Market Models

The Role of Behavioral Heterogeneity in Asset Pricing Models

In this chapter we empirically examine alternative models of behavioral heterogeneity in financial markets via moment-based estimation: the adaptive belief system (ABS) and a structural stochastic volatility (SSV) model. First, we discuss the empirical performance of ABS with two types of noise specification (i.e. additive and multiplicative) using S&P 500. In particular, we evaluate the validity of two model specifications from the simulated test distribution. Second, we consider two trading mechanisms of SSV and compare their empirical performance on the return volatility. The result of the formal test shows that the herding mechanism fits the data better than the model where the wealth mechanism is used, but the two models are not significantly different at 5% level.

5.1 Introduction

Agent-based models (ABM) are often used to account for behavioral heterogeneity and social interactions between investors. For example, De Long et al. (1990) as well as Brock and Hommes (1997, 1998) develop a framework for the asset pricing model in which price movements are endogenously determined by heterogeneous expectations. Similarly, several variants of behavioral heterogeneity in ABMs have been proposed during the last decade, which can describe the interactive trading processes observed in financial markets; see Gaunersdorfer (2000), Chiarella and He (2002), De Grauwe and Grimaldi (2006) and others. It has been shown that the model of behavioral heterogeneity is capable of depicting the relationship between price movements and trading mechanism.

It is generally assumed in the asset pricing model that the investor is aware of his or her investment objective: i.e. profit maximization by allocating the money to risky or risk-free assets. However, the main difference between ABMs and the traditional portfolio model lies in the fact that in the former market demand is primarily driven by a weighted average of different types of traders. Especially, the behavioral heterogeneity can be developed by applying the discrete choice theory to the trading behavior in the asset pricing models; also see Anderson et al. (1992). For example, the nonlinear dynamics of the adaptive belief system (ABS) are characterized by the behavioral rules of fundamentalist and chartist as well as their interactions;

the fundamentalist sticks to his or her belief on the market fundamental whereas the chartist takes up an adaptive position and seeks for opportunity from the price trend. In most cases, a simple decision rule, which is used in ABS, prevents the price series from being directed to an ever-increasing trajectory (or bubble). Accordingly, the interactive process of simple switching between the two investors can provide a potential explanation for the empirical regularities in financial data such as volatility clustering and fat-tailed distribution.

However, the non-linear mapping between the model structure and the reduced parameters hinders an empirical investigation of ABMs. For example, we cannot prove the global identification conditions for the parameters of ABMs, and their intractability does not allow us to investigate the convexity of the objective function. As a result, the empirical analyses of the ABMs pay attention to some restricted parameter space and delve into the modeling problems; Amilon (2008) used an auxiliary model to estimate a variant of ABS and found that the model with a realistic modeling of noise specification can not provide a good approximation to financial data; e.g. long range dependence in return volatility. Franke (2010) also discussed the importance of the noise term for the model dynamics; i.e. the original version of ABS cannot explain stylized facts known in financial literature.¹

In this chapter, first we review two asset pricing models with behavioral heterogeneity: ABS and a structural stochastic volatility (SSV) model. In the former, we discuss the importance of two types of noise specification for evaluating the empirical performance; i.e. additive and multiplicative. In the latter, we examine the role of trading mechanisms in the model dynamics (see also Franke and Westerhoff (2011a, 2012)[FW hereafter]). To see this, we compare the empirical performance of two trading mechanisms: namely, the investors can either (1) continue with a profitable strategy based on their past performance (wealth) or (2) follow the opinion of successful peer groups (herding). Second, we propose the simulated method of moments to estimate the parameters of ABS and SSV with the data of S&P 500 index. Since behavioral heterogeneity has a highly non-linear impact on the macroscopic variable, e.g. the market price, we can not find a simple closed form solution for these models. Therefore we cannot derive analytical moment equations of the system, but must approximate them based on simulations; see also Gilli and Winker (2003). Concerning the moment conditions in our empirical applications, we utilize four features of stylized facts known in financial literature following Franke and Westerhoff (2012): the absence of autocorrelations in the raw returns, fat-tailed distribution, volatility clustering, and long memory.

Since the non-linear mapping from the parameters of ABS to the objective function gives rise to multiple local minima during the estimation, we restrict our em-

¹To cope with identification problems in the non-linear and stochastic models, we can fix some of the model parameters and evaluate the empirical performance based on the stylized facts known in empirical literature; e.g. model validation. Moreover, we can calibrate some of key parameters by means of additional (micro-level) data sets, which are widely accepted by academics or researchers; see also the validation methodology by Winker et al. (2007) and the special issue of *Computational Economics* (2007) edited by Fagiolo et al. (2007) regarding empirical applications of ABMs.

empirical analysis to the parameters, which are investigated by Gaunersdorfer and Hommes (2007)[GH hereafter]. Therefore this study does not provide an elaborate analysis of the parameter estimation, but examines the effects of the noise term on the model dynamics. The results of the empirical application show that the baseline model of ABS cannot reproduce all stylized facts of financial data; i.e. the absence of long range dependence in absolute returns. Accordingly, the model with multiplicative noise has a poor performance on the financial data while the other model with additive noise is sensitive to the noise level for matching the empirical moments. Moreover, we discuss the importance of the finite size effects for analyzing the decision rules of ABS. The simulation results show that the stochastic extension of the switching rules allows for a possible misperception of investors in the deterministic dynamics.

Concerning the role of the trading mechanism in the price dynamics, we examine the empirical performance of SSV using a formal test. For instance, Franke and Westerhoff (2011b) investigate the fit of several variants of SSV and rank them based on their ability to match the chosen moment conditions using simulations (i.e. joint moment coverage ratio). In order to make the model comparison more explicit, we evaluate the fit of two competing specifications in SSV according to the simulated test distribution. In particular, we apply the formal test of Hnatkovska, Marmer and Tang (2012) [HMT hereafter] and examine a significant difference between the two trading mechanisms in fitting the data. We find that the herding mechanism has a better goodness-of-fit to the data than the model dynamics where the wealth mechanism is used, but the results reveal no statistically significant difference between them at 5% level.

The chapter proceeds as follows. Section 2 discusses the basic dynamic properties of the asset pricing model with behavioral heterogeneity; the effects of noise specification are investigated in ABS while the role of trading mechanisms is empirically examined in SSV. Section 3 explains the estimation methodology, and section 4 presents the empirical results as well as the formal test of SSV. Finally, section 5 concludes. Appendix collects all relevant details of this study.

5.2 The asset pricing models with behavioral heterogeneity

This section reviews a theoretical framework for the asset pricing model with behavioral heterogeneity. First, we discuss the role of heterogeneous trading rules in the price dynamics of ABS and examine the effects of noise specification on the price movements: additive and multiplicative shocks. Second, we investigate the effect of behavioral heterogeneity in terms of two trading mechanisms of SSV: herding and wealth.

5.2.1 The adaptive belief system

We examine the portfolio choice problem for the investor who manages the risk involved in asset markets using mean variance optimization. Note here that the model with the traditional portfolio theory has been extended allowing for heterogeneous expectations. There are two types of investment strategy in financial markets; i.e. fundamentalism and chartism. These strategies follow simple behavioral rules:

$$E_{1,t}[p_{t+1}] \equiv p_{1,t+1}^e = p^* + \nu(p_{t-1} - p^*), \quad 0 \leq \nu \leq 1 \quad (5.1)$$

$$E_{2,t}[p_{t+1}] \equiv p_{2,t+1}^e = p^* + g(p_{t-1} - p_{t-2}), \quad g \geq 0 \quad (5.2)$$

The fundamentalists (type 1) believe that tomorrow's price will move to the fundamental price p^* by the adjustment speed of ν ; if $p_{t-1} < p^*$, for example, then the fundamentalists expect that this difference will be covered by the market clearing, but with a time lag of the adjustment speed in the market.² On the other side, the chartists (type 2) make use of information about the observed price trend and continuously update this investment opportunity with the adjustment speed of g .

In this model, we assume that the investors allocate their wealth in two ways, namely either to keep their money in a risk-free asset or to invest it into a risky asset. Because of this simplification, the total wealth of two traders (type $h=1, 2$) in period $t+1$ can be expressed as the current wealth and the expected profits of the investment:

$$W_{h,t+1} = (1+r)W_{h,t} + \{p_{t+1} + \bar{y} - (1+r)p_t\} \cdot z_{h,t} \quad (5.3)$$

where the state variable $z_{h,t}$ refers to the market demand for the risky asset from a trader type h . The variables \bar{y} and r denote a constant dividend and interest rate. For the sake of simplicity, we use the notations R and R_{t+1} for $1+r$ and $p_{t+1} + \bar{y} - (1+r)p_t$, respectively.

It is assumed here that investors maximize the expected return, but minimize the risks from their investment. This statement of the problem can be formulated as:

$$\Phi \equiv \max_{z_{h,t}} E_{h,t}(W_{h,t+1}) - \frac{a}{2} \cdot V_{h,t}(W_{h,t+1}), \quad h = 1, 2 \quad (5.4)$$

²Without loss of generality, we assume that the fundamental price is constant over time. This implies that the deviation of market price from the fundamentals is important to generate return volatility in this model.

where the parameter a measures risk aversion ($a > 0$). The optimality condition of Equation (5.4) can be expressed as its first derivative. We set it equal to zero with respect to the market demand:

$$\frac{\partial \Phi}{\partial z_{h,t}} = E_{h,t}(R_{t+1}) - a \cdot \overline{Var}(R_{t+1})z_{h,t} \stackrel{!}{=} 0 \quad (5.5)$$

Once the market demand and supply are cleared, we can arrive at the myopic mean variance optimization for heterogeneous agents:

$$z_{h,t} = \frac{E_{h,t}(R_{t+1})}{a \cdot \overline{V}_{h,t}(R_{t+1})} \quad (5.6)$$

For the sake of simplicity, we assume that investors have a constant belief of variance ($\overline{V}_{h,t} \equiv \sigma^2$). Since the trading strategies influence the wealth dynamics through Equation (5.3), we can simply indicate the evolutionary fitness measure by using accumulation of the profits:

$$\begin{aligned} U_{h,t} &= (1+r)z_{h,t-1} + \eta U_{h,t-1} \\ &= \frac{R}{a\sigma^2} \cdot \{p_{h,t}^e + \bar{y} - (1+r)p_{t-1}\} + \eta U_{h,t-1}, \quad h = 1, 2 \end{aligned} \quad (5.7)$$

where η is the memory parameter measuring how fast the past fitness is discounted with the selected market strategies.

Note here that the investors update their utility and evaluate the chosen trading strategy until the expected profits do not differ from the market realization. In other words, the investors assess their preferences ordering between the two trading strategies, which are determined by the utility function of their fitness values. Especially the intermediate dynamics of ABS are associated with the switching process between two investors; i.e. the discrete choice theory. In this framework, it can be seen that the observed choice between fundamentalism and chartism is made according to the relative advantage of utility from the fitness measure.³

$$\bar{n}_{h,t} = \frac{\exp(\beta U_{h,t-1})}{\exp(\beta U_{1,t-1}) + \exp(\beta U_{2,t-1})}, \quad h = 1, 2 \quad (5.8)$$

³The discrete choice framework for the group dynamics is commonly associated with Gibbs distribution when the infinite number of agents is assumed in the system. See also Manski and McFadden (1981) for details.

where the parameter β denotes the intensity of choice. The resulting evolutionary fitness and the choice probability shape the trajectory of the price dynamics. This implies that the price movements can be considered as being a direct result of the interactive adjustment processes between the fundamentalists and the chartists.⁴

We complete the updating scheme by assuming that the fraction of traders evolve according to the following rule:

$$n_{2,t} = \bar{n}_{2,t} \cdot \left[\exp \left\{ \frac{-(p_{t-1} - p^*)^2}{\alpha} \right\} \right] \quad (5.9)$$

It can be seen in Equation (5.9) that the updating capacity of the chartist is controlled by the parameter α , which measures the sensitivity to the observed deviation from the fundamentals. Finally, we arrive at the price equation in the model with the heterogeneous market demand: the interactive process gives rise to market price movements. Therefore the equilibrium price is reflected in the demand of the heterogeneous traders after adjusting their market shares:

$$\hat{p}_t := \frac{1}{R} \{ n_{1,t} \cdot E_{1,t}(p_{t+1} + \bar{y}) + n_{2,t} \cdot E_{2,t}(p_{t+1} + \bar{y}) \} \quad (5.10)$$

In sum, the four lagged prices (p_{t-j} , $1 \leq j \leq 4$) and two lagged fitnesses ($U_{h,t-1}$, $U_{h,t-2}$) affect the current price in the model. Concerning the computation of the model, we can simulate the model by using the initial four price values. Then the model can sequentially generate the price series from Equation (5.10). During the intermediate steps, the heterogeneous trading rules have a direct impact on their realized profits. The performance of strategy especially has influence on the evolutionary part of the model in which the price dynamics are endogenously governed by the market fractions of heterogeneous traders.

Table 5.1: Parameter values for simulation

ν	g	α	β	r	a	σ	η	ρ	\bar{y}	P^*
1	1.9	1800	2	0.001	1	1	0.99	10	1	1000

Accordingly, the model dynamics can reproduce a periodic ups and downs in the price path; i.e., see the limit cycle in Figure 5.1. Note here that the simulated series are generated using the parameter values in Table 5.1. Indeed, whenever the chartists are dominant in the market, the model predicts that the price will be deviated from the fundamentals. Since the behavior of chartists will be influenced by this

⁴However, the model dynamics become unstable when this parameter of the switching rules increases more than two ($\beta > 2$); e.g. see Gaunersdorfer (2000) and Gaunersdorfer et al. (2008) for the role of this intensity of choice parameter in the system bifurcation as well as Goldbaum and Mizrach (2008) for estimation.

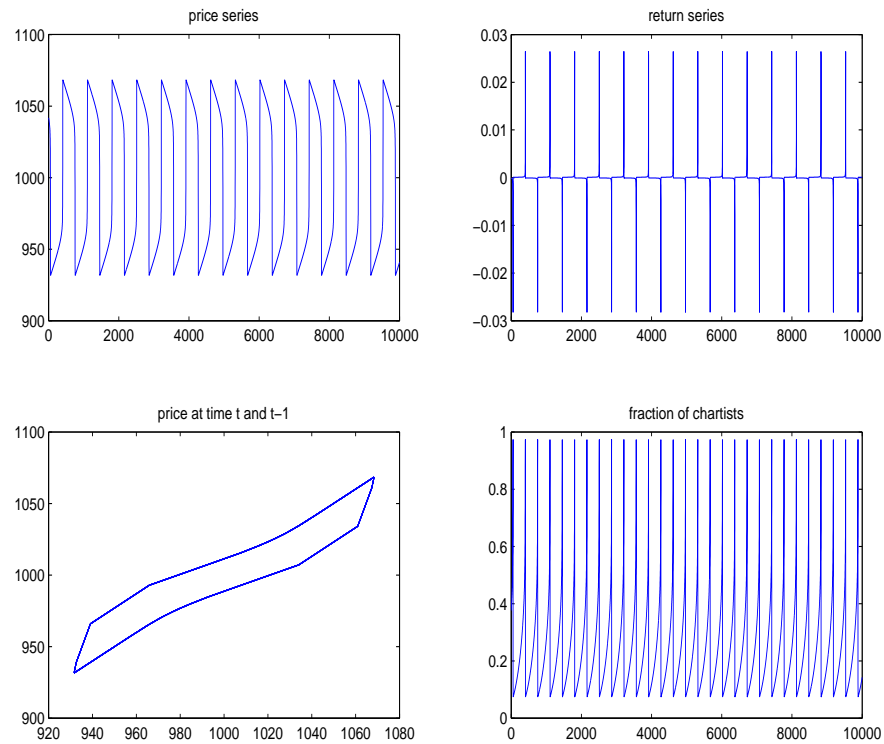


Figure 5.1: Deterministic structure of ABS

observed deviation (penalty term), however, this will suppress the market activity of the chartists (e.g. to what extent the price can be deviated from the fundamentals, $\exp\left\{\frac{-(p_{t-1}-p^*)^2}{\alpha}\right\}$). This ensures that the updating process can place a limit on the deviation from the fundamentals. On the other hand, if the chartist strategy is not profitable anymore, then the investors are likely to take up the opposite position (fundamentalism). Overall, it can be seen that the price fluctuations in the model are considered results of the interactive processes between the investors.

Noise specification in ABS

In the current study, we consider two types of the noise term for simulating the price equation of ABS: additive and multiplicative. The effect of noise specification has been extensively studied in the field of stochastic differential equations; also see Longtin (2003) for the use of noise term in Langevin equations. The market equilibrium price with additive noise can be defined as:

$$p_t = \hat{p}_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \rho_1) \quad (5.11)$$

where ρ_1 measures the magnitude of additive noise over time. \hat{p} denotes the deterministic part of the price dynamics in ABS.

However, we observe a high noise-to-price ratio in the model dynamics when the simulated price level is moderately low; if $p_t = 50$ and $\rho = 10$, then the noise-to-price ratio ($\frac{\varepsilon_t}{p_t}$) becomes 20%. When the noise drastically change the system, for example, it can be seen that the model allows for multiple equilibria in the price trajectory. The intermittent switching processes are sensitive to the initial price values used in simulations. As an alternative route to randomness in the model dynamics, we can propose the market price equation with multiplicative noise, where the noise term and the lagged price have the multiplicative effects on the current price level; see Amilon (2008) and Franke (2010).

$$p_t = \hat{p}_t + p_{t-1} \cdot \varepsilon_t, \quad \varepsilon_t \sim N(0, \rho_2) \quad (5.12)$$

where ρ_2 measures the magnitude of multiplicative noise over time.

The parameter values in Table 5.1 are used for simulating return series; e.g. additive ($\rho = 10$) and multiplicative ($\rho = 0.01$) shocks in Equation (5.11) and (5.12), respectively. Figure 5.2 illustrates that the returns with the multiplicative shocks (right panel) show less dependence in volatility than the other one (left panel). Moreover, it can be seen that the model with additive noise cannot predict the volatility clustering in financial data, because the simulated data exhibit strong time dependence without having a distinct impact of chartism on the model dynamics (see the period between 3,500 and 4,500 in the left panel).

Note here that we can easily simulate the decision rules of two investors, but the system can undergo a bifurcation with some parameter combinations. This can allow for a negative price trajectory in simulations; i.e. $p_t < 0$. Because of this, we do not attempt to estimate all parameters in the model, but analyze the relevance of noise specification in the model dynamics; i.e. to what extent the specification of the noise term affects the model dynamics in fitting real data. It will be discussed later in detail.

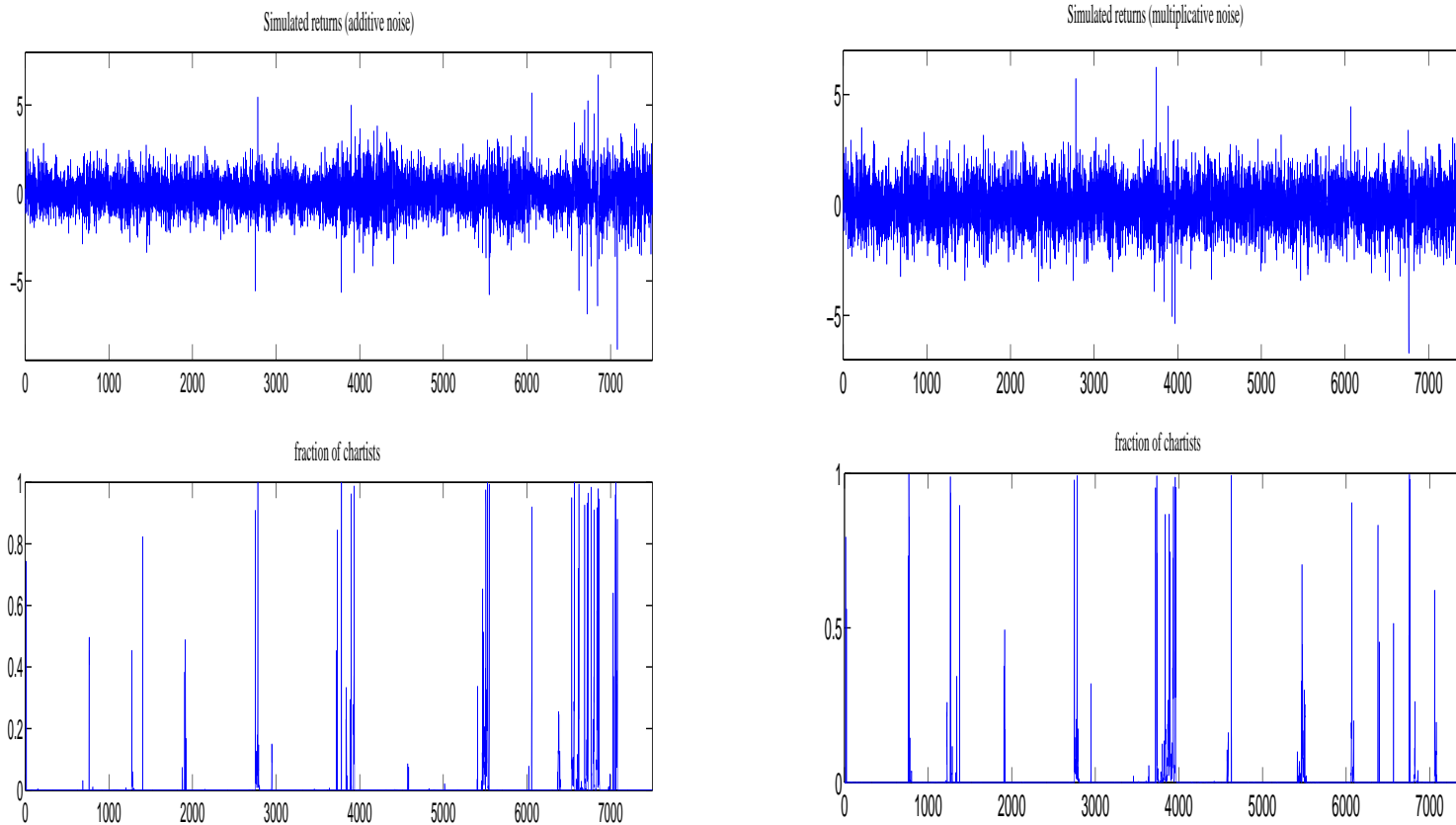


Figure 5.2: Return series of ABS with additive (left) and multiplicative (right) noise.

We use the parameter values of Table 5.2 for simulations. The values for the noise term ρ are 10 (additive) and 0.01 (multiplicative) respectively.

5.2.2 A structural stochastic volatility model: herding versus wealth

In this section we discuss the herding and wealth mechanism of the SSV model. At the beginning, we distinguish two types of investors: fundamentalists and chartists. Their trading behaviors of the mean reversion and the trend following strategy serve a market demand for risky assets. The behavioral heterogeneity developed in this model is similar with ABS, but the main difference lies in the specification of demand shocks for the two investors:

$$d_t^f = \phi(p^* - p_t) + \varepsilon_t^f, \quad \varepsilon_t^f \sim N(0, \sigma_f^2) \quad (5.13)$$

$$d_t^c = \chi(p_t - p_{t-1}) + \varepsilon_t^c, \quad \varepsilon_t^c \sim N(0, \sigma_c^2) \quad (5.14)$$

where the state variables d_t^f and d_t^c denote the demand of average fundamentalists and chartists. The parameters ϕ and χ measure the response to price changes from the fundamentals (p^*) and the previous price (p_{t-1}), respectively. The demand shocks are represented as two independent and normally distributed random variables ε_t^f and ε_t^c for the two investors.

Concerning the system size in the model, we assume that the market is populated by the total number of $2N$ fundamentalists (n_t^f) and chartists (n_t^c). Note here that we regard the majority index of the fundamentalists as a proxy for the market regimes: fundamentalism (" $x = +1$ ") or chartism (" $x = -1$ ").

$$x_t := \frac{n_t^f - n_t^c}{2N} \quad (5.15)$$

Accordingly, the price dynamics are driven by the market demands from two types of investors. Indeed, it can be seen that the interactive market processes prevail when the market is in disequilibrium. The adjustment equation for the price at period $t + 1$ can be set up from Equations (5.13) to (5.15) as follows:

$$p_{t+1} = p_t + \frac{\mu}{2} \cdot [(1 + x_t) \cdot \phi \cdot (p^* - p_t) + (1 - x_t) \cdot \chi \cdot (p_t - p_{t-1}) + \varepsilon_t] \quad (5.16)$$

$$\varepsilon_t \sim N(0, \sigma_t^2), \quad \sigma_t^2 = \frac{(1 + x_t)^2 \sigma_f^2 + (1 - x_t)^2 \sigma_c^2}{2}$$

As a rule, the excess demand can be adjusted by a market maker, who can increase or decrease the market price with a constant proportionality factor $\mu > 0$. Under these circumstances, the total demand shocks can be summarized by two normal random distribution with mean zero and two single variances. Hence, the model predicts that the price volatility is driven by market structure in the economy, and

the noise term σ_t^2 in Equation (5.16) changes according to heterogeneous demand shocks. Especially one obtains the current demand shock from the time-varying share of the fundamentalists $(1 + x_t) \cdot \sigma_f$ and the chartists $(1 - x_t) \cdot \sigma_c$.⁵

Now we discuss the dynamics of SSV in terms of behavioral heterogeneity. In the same way of ABS, the switching mechanism based on the discrete choice theory provides a description of the group dynamics in SSV. Then the market shares are solely explained by the following equation:

$$n_t^s = \frac{\exp(\beta u_{t-1}^s)}{\exp(\beta u_{t-1}^f) + \exp(\beta u_{t-1}^c)}, \quad s = f, c \quad (5.17)$$

where β denotes the intensity of choice parameter. To be more specific, we divide Equation (5.17) by the observed market utility of the fundamentalists $\exp(\beta u_{t-1}^f)$. Therefore the equation of the market share can be transformed into compact form, from which we easily examine the share of fundamentalists regarding their utility advantage over the chartists:

$$n_t^f = \frac{1}{1 + \exp\{-\beta(u_{t-1}^f - u_{t-1}^c)\}} \quad (5.18)$$

$$n_t^c = 1 - n_t^f$$

For the sake of simplicity, we redefine the discrepancy between utilities for two groups (i.e. $u_{t-1}^f - u_{t-1}^c$) as an attractive index a_{t-1} . Since the past capital gains give rise to the payoff difference between the fundamentalists (u^f) and the chartists (u^c), the observed utility advantage, which is denoted by the previous attractive index, has influence on the realization of the group dynamics. Turning to the specification for the attractive index, we examine the variants of the trading behavior, i.e. the herding and wealth mechanism.

- Discrete choice with herding mechanism

In order to establish a connection between the group behavior and return volatility, we define the attractive index a_t as a payoff equation with the trading strategy of mean reversion and the majority index (Equation (5.19)). By incorporating information about the market sentiment (i.e. x_t) into the current payoff of investors, the herding mechanism predicts the price dynamics in terms of the socio-economic configuration of the market.

⁵The transformation of the majority index by adding/subtracting unity results in the relative share of fundamentalists/chartists in each group; i.e. $1 + x_t = \frac{n_t^f}{N}$, $1 - x_t = \frac{n_t^c}{N}$.

$$a_t(x_t, p_t) := \alpha_0 + \alpha_x x_t + \alpha_d (p_t - p^*)^2 \quad (5.19)$$

where the coefficient α_0 measures a predisposition of traders to the fundamentals. The parameters α_x and α_d denote a group pressure and the influence of misalignment, respectively ($\alpha_x, \alpha_d > 0$).

- Discrete choice with wealth mechanism

In the wealth mechanism, we assume that investors compare their profits with the other group; this payoff difference will have influence on their investment strategy. The primary difference between the herding and wealth mechanism lies in the specification of the attractive index; the observed discrepancy of capital gains between the two investors controls an endogenous switching process in the model of the wealth mechanism.

$$\begin{aligned} g_t^s &= [\exp(p_t) - \exp(p_{t-1})] \cdot d_{t-2}^s \\ w_t^s &= \eta w_{t-1}^s + (1 - \eta) g_t^s, \quad s = f, c \end{aligned} \quad (5.20)$$

$$a_t(x_t, p_t) := \alpha_w (w_t^f - w_t^c)$$

where we express the current wealth as a weighted sum of the previous wealth (w_{t-1}) and current profits (g_t). The coefficient η measures their memory on the previous price ($0 \leq \eta \leq 1$), while the parameter α_w denotes the adjustment speed of the pay-off strategies in the switching process ($\alpha_w > 0$).

5.3 The estimation methodology

In this study, we estimate the parameters of ABS and SSV via the simulated method of moments (SMM) estimator. Since the decision rules from the discrete choice theory make these models highly non-linear, we cannot easily obtain the analytical moment equations of a system. In this case, the SMM estimation can replace the moment conditions by approximations based on simulations. We can obtain the parameter estimates by minimizing the distance between the model-generated and empirical moments in the objective function. For the comparison of two competing models, the objective function can be defined as:

$$J_I(\theta) = \min_{\theta^I \in \Theta} [\widehat{m}_T - m^I(\theta^I)]' \widehat{W} [\widehat{m}_T - m^I(\theta^I)], \quad I = A, B \quad (5.21)$$

where we distinguish the estimated objective function values of the two models using the notation I ; i.e. $I = A, B$. m^I denotes a vector of the model-generated moments while \hat{m} is an estimate of the true moments m_0 . In particular, the fit of the two model specifications to the data is compared using the formal test of HMT.

As it is discussed above, the switching process based on the discrete choice theory affects the dynamics of ABS and SSV in a non-linear way. Therefore we approximate the moment conditions of these models by simulations; i.e. $m^I(\theta^I) = \frac{1}{\psi T} \sum_{t=1}^{\psi T} m_t$. We denote by ψ a simulation size, which is set to 100.

The weight matrix is constructed using the moving block bootstrap method. We denote by m^b a vector of the bootstrapped sample moments. B is the number of block bootstrap sampling. \bar{m} is the mean value of the bootstrapped moments.

$$\hat{\Omega}_{BB} := \text{Var}(m^b) = \frac{1}{B} \sum_{b=1}^B (m^b - \bar{m})(m^b - \bar{m})' \quad (5.22)$$

As a result, we can compute the weight matrix by taking the inverse of the variance-covariance matrix from 1,000 moving block bootstrapped sample: $\widehat{W} = \hat{\Omega}_{BB}^{-1}$.

Turning to the block length, we use a fixed window of 250 days.⁶ To make the test procedure of HMT (Hnatkowska, Marmer, Tang (2011)) [HMT-sup henceforth] operational, we estimate the variations in the moment estimates and their covariance structure by using the double block bootstrap method; see appendix D.

Under the standard regularity conditions, $\sqrt{T}(\theta_T - \theta_0)$ converges in distribution as $T \rightarrow \infty$ to a normal random vector with mean zero and covariance Λ_ψ (Lee and Ingram (1991), Duffie and Singleton (1993)):

$$\Lambda_\psi = [(D_\psi W D_\psi')^{-1}] D_\psi' W (1 + 1/\psi) \Omega W D_\psi [(D_\psi W D_\psi')^{-1}]' \quad (5.23)$$

where D_ψ is an expected partial derivative matrix of the moments with respect to simulations used in the model. This can be numerically computed at optimum of the parameter estimates ($D_\psi \equiv E \left[\frac{\partial m(\theta)}{\partial \theta} \right]$).

⁶If researchers believe in long range dependence in their data, then they can increase the window length for estimating the covariance structure of the moments in the bootstrap method. See also a performance comparison of several double blocks-of-blocks procedures by Bühlmann (2002), Horowitz et al. (2006), Lee and Lai (2009) and others. In our study we use the one-year window, because we cannot find any significantly positive autocorrelation in S&P 500 index after one year; see also Winker and Jeleskovic (2007).

5.4 Empirical application

In this section we first present the data and choice of moment conditions for our empirical application. Second, we discuss the parameter estimates of both ABS and SSV. Finally, we conduct a formal test to decide whether the fit of the two trading mechanisms of SSV is significantly different.

5.4.1 Data and choice of moments

For the empirical application, we use the historical data of S&P 500 index. The daily data range from Jan.02.1980 to Sep.29.2006. The sample size of returns is 6750. We use the 10 moment conditions, which are characterized by four features of stylized facts known in financial literature. For example, Pagan (1996) studied a set of stylized facts on financial data, while Winker and Jeleskovic (2006, 2007) discussed the unconditional and time-dependent properties of foreign exchange rate series.

Table 5.2: Empirical and bootstrapped moment estimates of S&P 500 index

label		empirical	moving BB
m_1	mean	0.0427	0.0421
m_2	variance	1.0628	1.0706
m_3	$\text{corr}(r_t, r_{t+1})$	0.0196	0.0191
m_4	$\text{corr}(r_t , r_{t+1})$	0.1825	0.1746
m_5	$\text{corr}(r_t , r_{t+5})$	0.2131	0.2041
m_6	$\text{corr}(r_t , r_{t+10})$	0.1531	0.1455
m_7	$\text{corr}(r_t , r_{t+20})$	0.1264	0.1138
m_8	$\text{corr}(r_t , r_{t+50})$	0.1070	0.0850
m_9	$\text{corr}(r_t , r_{t+100})$	0.0703	0.0406
m_{10}	5-10% right tail of Hill estimator	3.2722	3.2981
test criterion		95%: 19.07	99%: 25.44

Note: Moving BB means the moving block bootstrapped data. The summary statistics are computed using 1,000 bootstrapped resampling.

Table 5.2 presents the empirical and block bootstrapped moment estimates of the S&P 500 index. The mean (m_1) and variance (m_2) capture the general shape of distribution of the data. The 1st lag of autocorrelation of raw returns (m_3) is chosen to cope with the empirical fact that returns are not serially correlated; if the 1st lag appears to have no correlation, we are confident that raw returns do not show significant dependence over the subsequent lags. The autocorrelations of absolute returns ($m_4, m_5, m_6, m_7, m_8, m_9$) measure the short- and long-term dependence of volatility in returns. Finally, the Hill estimator (m_{10}) estimates the degree of the fat-tailed distribution in the data.

Figure 5.3 depicts a kernel estimate on the distribution of the objective function, which is obtained from 1,000 bootstrapped samples. The 95% and 95% criteria of this distribution are 19.07 and 25.44, respectively. These threshold values will be used as test statistics for evaluating the misspecification of the model in SMM; if the objective function value exceeds this criterion, we conclude that the model cannot match the data along the chosen moment conditions in this study.

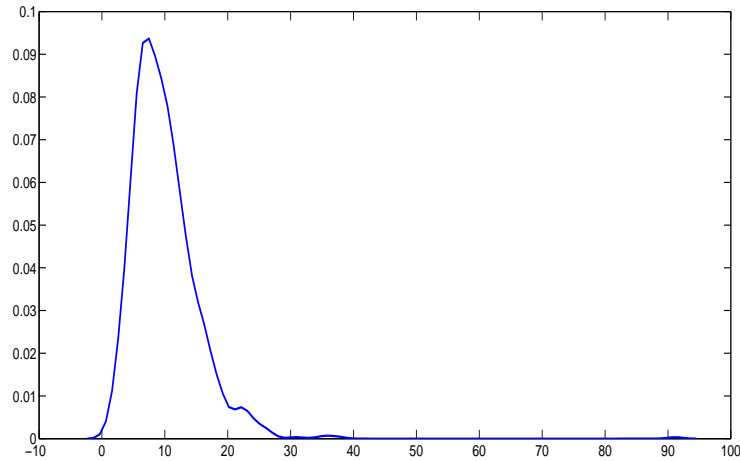


Figure 5.3: Distribution of the bootstrapped objective function values

5.4.2 Estimation results of ABS

In ABS, we focus on the stable region of the model dynamics for the parameter estimation; since the non-linear dynamics allow for a bifurcation in the system with respect to the model parameters (especially β), it can be seen that the mapping between the parameters θ and the model-generated moments $m(\theta)$ is highly non-linear. Accordingly, the objective function $J(\theta)$ does not have a unique minimum during the estimation.⁷ Therefore we restrict our analysis to the region in parameter space where the model dynamics are stable; i.e. we constrain the parameter estimates near the values, which are reported in GH for the model simulation. In addition, we implement the random search method to find a best minimum value for the parameters of ABS.

To cope with identification problems in the non-linear dynamics, we calibrate some parameters in the model; this can reduce the bias of parameter estimates. First, the parameter ν is set to unity where we can examine the case of the strong volatility clustering region for 'Efficient Market Hypothesis' believers; since all market information is already incorporated into the price, the fundamentalists expect

⁷In this case, econometric analysis should cope with multiple local minima in the objective function. As long as the dynamics of ABS depend on non-linear representation, the moment-based estimation could get trapped in local minima during the optimization process.

that the future price will follow the lagged one ($p_{1,t+1}^e = p_{t-1}$). Second, we set the risk aversion and the constant belief on the variance to unity, respectively; i.e. $a = 1$, $\sigma = 1$. The other behavioral parameters in the discrete choice framework are calibrated following the previous simulation results of GH: $\alpha = 1800$, $\beta = 2$. Hence, we have now three parameters to be estimated in the model: adjustment speed of chartists g , memory parameter η , noise term ρ .

Table 5.3: Parameter estimates for ABS

label	additive		multiplicative	
g	5.498	(1.287)	6.488	(1.491)
η	0.981	(0.254)	0.378	(0.154)
ρ	8.500	(0.431)	0.008	(0.001)
$J(\theta)$	107.07		133.24	

The other parameters are calibrated: $\nu = 1$, $a = 1$, $\sigma = 1$, $\alpha = 1800$, $\beta = 2$. (): standard error.

Table 5.3 reports the estimates of two specifications in the model.⁸ Indeed, the estimated values for g do not differ too much between the models with additive and multiplicative noise. Especially, this empirical result shows that the fit of the model relates to the return volatility; the higher the parameter value for g , the more volatile the price movements. However, the model with the additive noise attains a higher value for memory parameter than the other (η : 0.98 (additive) > 0.38 (multiplicative)). Since the role of noise term in the two models is not same, we cannot make a direct comparison between the estimated values for the shocks. However, it can be seen that the magnitude of the shocks for both models is similar; since the fundamental price is set to 1,000 for the model simulation, we can arrive at $\bar{\rho} = \frac{\rho}{p^*} = 8.5/1000 = 0.0085$.

Table 5.4: Simulated moments of ABS with additive and multiplicative noise

Label	additive	multiplicative	Label	additive	multiplicative
m_1	-0.0004	-0.0002	m_6	0.0444	0.0112
m_2	1.0255	1.0938	m_7	0.0401	0.0042
m_3	0.0473	0.0810	m_8	0.0414	0.0024
m_4	0.1017	0.0790	m_9	0.0342	0.0008
m_5	0.0465	0.0157	m_{10}	3.8535	3.7292

⁸Since the model estimation is obtained using a random search method, the point estimates of the model do not indicate the smoothed minimum value in the objective function. Therefore we do not apply the asymptotic normality of the parameter estimates from Equation (5.23). Alternatively we compute the standard error of the parameter estimates by using 100 replications of SMM with different random seeds.

More generally, the empirical results can be summarized as follows:

- From this empirical application, we found that the model dynamics of ABS are misspecified, since the objective function values of the two models exceed the simulated test criterion ($\widehat{J}(\theta)$: $107.07, 133.24 > 25.44$).
- Concerning the minimum values of the objective function, we do not have much confidence in the uniqueness of the parameter estimates, because different parameter estimates are often found when we change starting values in optimization. This violates the assumption 2.4 of HMT, especially the uniqueness of parameter estimates. Therefore we do not compare the fit of two models of ABS using the formal test.

Alternatively we compare the fit of the two models from the model-generated moments. Table 5.4 shows the simulated moment conditions from the two specifications. For example, they do not have any difficulty to match the mean and variance of the data (m_1, m_2). But these models fail to approximate the characteristics of the data, especially the time dependence of return volatility; e.g. see m_4, m_5, m_6, m_7, m_8 . The empirical performance for matching these moments in the model with multiplicative noise is worse than the other model. By applying the Hill estimator, we found that the model-generated moments contain slightly less extreme values than the empirical ones; i.e. 3.85 (additive), 3.73 (multiplicative) $>$ 3.27 (empirical).

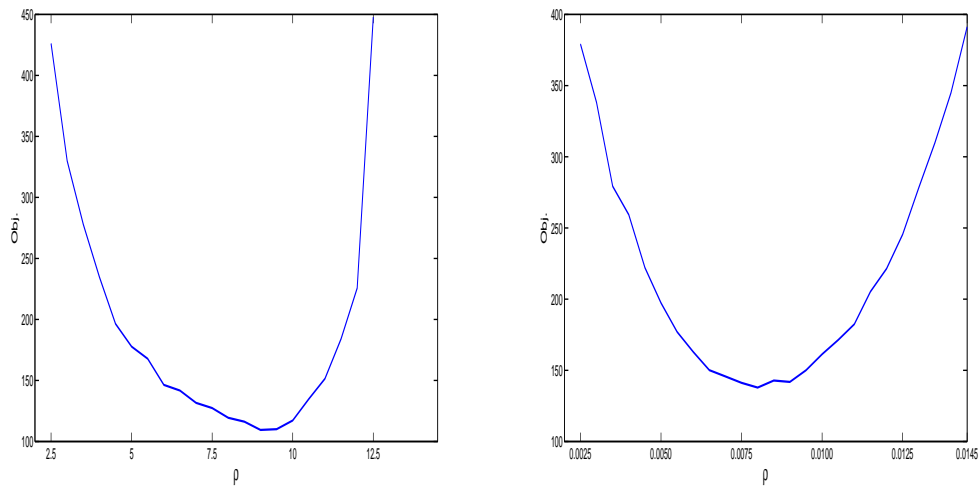


Figure 5.4: The effect of noise specification on objective function: additive (left) and multiplicative (right).

To examine to what extent the level of the noise term plays a role in approximating the data, we investigate the curvature of the objective function with respect

to the scale of the shock. For this computation, we set the other parameters (g and η) to the values in Table 5.3, and change the degree of the noise term. Then we calculate the objective function based on the noise level and depict their variations. In Figure 5.4, the left panel draws the objective function of the model with additive noise while the right panel accounts of the case with multiplicative noise. Especially, when the scale of the additive noise increases, we obtain a much higher value for the objective function, i.e. a bad fit of the model to the data. This graphical comparison between the two models shows that the model with additive noise is more sensitive to the level of noise assumed in the model than the other. Overall, the results indicate that the noise level is empirically important for simulating the dynamics of ABS.

5.4.3 Estimation results of SSV and model comparison

In this section, we discuss the empirical results of SSV and conduct a formal model comparison between two trading mechanisms. The level parameters of SSV are calibrated; i.e. the adjustment parameter for fundamentalists ϕ and chartists χ , the proportionality factor of market maker μ . In addition, we set the intensity of choice parameter β to unity. In the case of the herding mechanism, these parameters are set to the following values: $(\beta, \phi, \chi, \mu) = (1.0, 0.12, 1.5, 0.01)$. For the case of the wealth mechanism, they are set to the following values: $(\beta, \phi, \chi, \mu) = (1.0, 1.0, 1.0, 0.01)$.⁹

We use SMM to estimate the parameters of the two competing specifications in SSV. The parameter estimates are reported in Table 5.5. The single comparison of these parameter estimates is not much informative, because the two models use different specifications for the attractive index. But the estimated values for the demand shocks are noteworthy to mention here. For example, the estimation result of the herding mechanism indicates that the demand shock of the chartist is much higher than the wealth mechanism ($\hat{\sigma}_c$: 2.98 (herding) > 1.96 (wealth)), while the estimated demand shock of the fundamentalists shows a slight difference between the two models ($\hat{\sigma}_f$: 0.79 (herding), 0.68 (wealth)). This implies that the model approximation of the herding to the data depends largely on the demand shock of the chartists, whereas the dynamics of the wealth mechanism are more or less driven by the adjustment process implied by the observed discrepancy between the two investors ($\hat{\alpha}_w$).

⁹The level parameters are chosen according to FW. If we change these parameters, SMM would arrive at some inferior values of the objective function. Note here that this chapter demonstrates how the formal test can be applied to this model, but not an elaborate estimation.

Table 5.5: Parameter estimates of herding and wealth

label	herding	wealth
$\hat{\alpha}_0$	-0.417 (-0.640 ~ -0.195)	-
$\hat{\alpha}_x$	1.773 (1.280 ~ 2.267)	-
$\hat{\alpha}_d$	19.022 (11.128 ~ 26.916)	-
$\hat{\alpha}_w/100$	-	21.248 (0.000 ~ 99.129)
$\hat{\eta}$	-	0.986 (0.981 ~ 0.992)
$\hat{\sigma}_f$	0.788 (0.706 ~ 0.870)	0.684 (0.609 ~ 0.759)
$\hat{\sigma}_c$	2.983 (2.414 ~ 3.551)	1.957 (1.522 ~ 2.391)
$\hat{J}(\theta)$	26.11	49.85

Note. The brackets () indicate the 95% confidence intervals of the parameter estimates.

Now we evaluate the empirical performance of the two models by using a formal test. According to Vuong (1989), two models are overlapping if (i) they have common moment conditions and (ii) neither model is nested in the other. The SSV model with wealth and herding mechanism has the same underlying structure except for the specification of the attractive index. That is the main reason why we can consider the two specifications as being an overlapping case.¹⁰

Following HMT, we use two step sequential test. First, we compare the minimum value of objective function for the two models, which is obtained using SMM; i.e. herding: 26.11, wealth: 49.85, respectively. Then we measure the significant difference between the two models by subtracting these values (i.e. 23.74); see chapter two and three regarding the formal framework for the model comparison.

The critical value for the model comparison is 141.21 at 5% level; see the computational details for the simulated distribution in appendix C as well as Figure 5.8 in appendix E. The estimated difference between two objective function values does

¹⁰To make this classification more specific, we denote by $\theta^h = (\alpha_0, \alpha_x, \alpha_d)'$ and $\theta^w = (\eta, \alpha_w)'$ the parameter set of the herding and wealth mechanisms, respectively. First, the two models are not strictly non-nested, because from some combinations of the parameters (namely, $\theta^h = 0$ and $\theta^w = 0$), we can find that they have same moment conditions, i.e. $m^A(\theta^h) = m^B(\theta^w)$. In other words, the ability of the two models to match the characteristics of the data is potentially equivalent; i.e. the fit of two models to the true data generating process are equally accurate. Second, one model is not nested in the other, because the trading mechanism will have different influence on the switching mechanisms and produce the different moment conditions. This means that we have $m^A(\theta^h) \neq m^B(\theta^w)$ asymptotically if $\theta^h \neq 0$ and $\theta^w \neq 0$. Therefore we treat the two models as being an overlapping case.

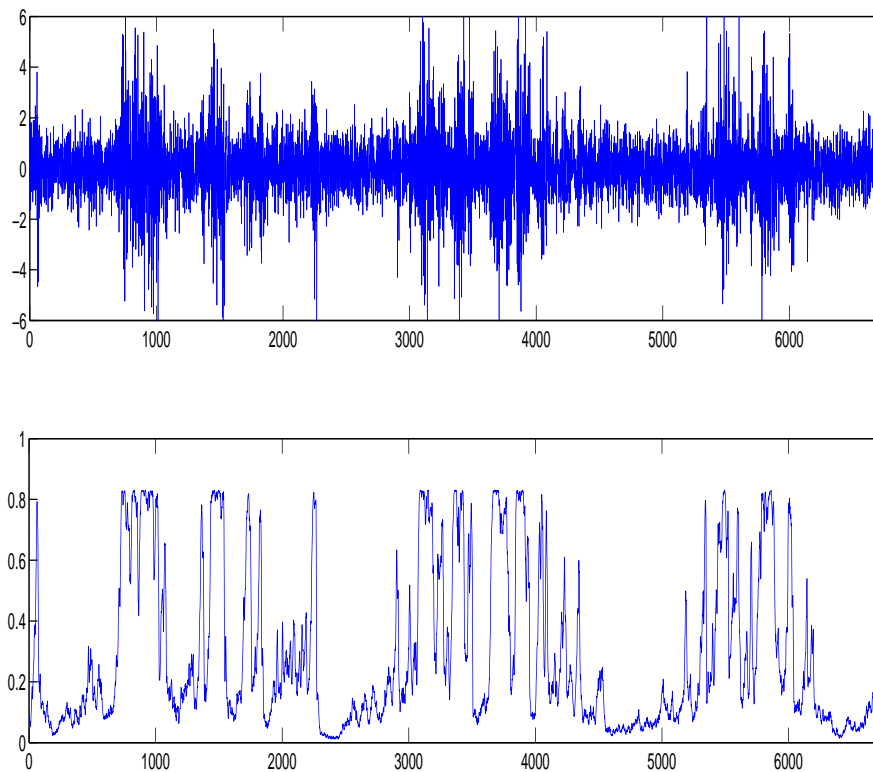


Figure 5.5: Returns (upper) and fraction of chartists (lower) from herding

not exceed the simulated critical value (i.e., $49.85 - 26.11 = 23.74$). Therefore we cannot reject the null hypothesis that the two models have an equal fit to the data; we do not proceed to the second step of the model comparison.

Table 5.6: Simulated moments of herding and wealth mechanisms

label	herding	wealth	label	herding	wealth
m_1	0.0000	0.0000	m_6	0.2076	0.1978
m_2	1.3061	1.3506	m_7	0.1715	0.1660
m_3	0.0077	0.0068	m_8	0.1051	0.1036
m_4	0.2394	0.2413	m_9	0.0563	0.0487
m_5	0.2261	0.2154	m_{10}	2.9115	3.2849

To judge the validity of our comparison results in detail, we investigate a summary statistic of the individual moment conditions generated by each model. It can be seen in Tables 5.2 and 5.6 that the moment matching from both herding and

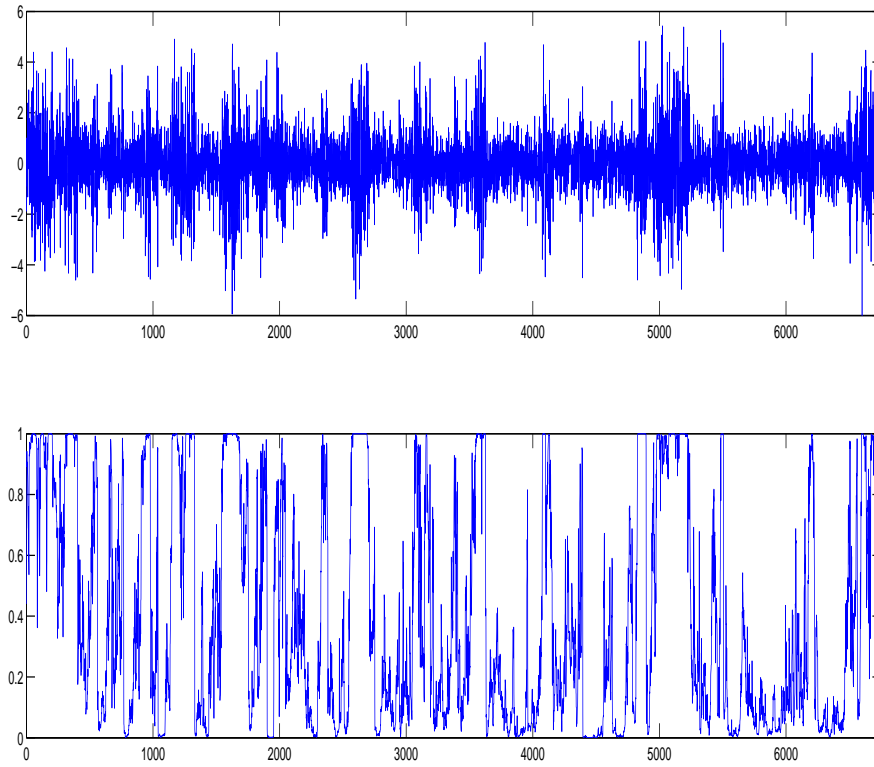


Figure 5.6: Returns (upper) and fraction of chartists (lower) from wealth

wealth is fairly good. Moreover, the numerical values for the estimated objective function indicate that the model-generated moments are not significantly different from the empirical ones. Indeed, the historical return volatility of S&P 500 can be explained by both of them; see Figure 5.5 and 5.6. We also point out that the inconclusive comparison results would be related to a sampling variance and the parameter uncertainty from the switching mechanism. In other words, the formal test of different trading mechanisms is sensitive to the amount of observations, i.e. the data set that contains a large number of extreme returns. Overall, the results show that the SSV model is moderately robust to the different switching mechanisms when matching the empirical features of the data. Put differently, the variations of the market fraction play a major role in approximating the stylized facts of financial volatility; the higher fraction of chartists, the higher volatility in returns.

5.5 Conclusion

In this chapter, we discussed the asset pricing models of behavioral heterogeneity and evaluated the empirical performance of ABS and SSV. In ABS, we examined the effects of noise specification on the model dynamics by comparing two types of the noise term. Since the empirical application of ABS violates standard assumptions in moment-based estimation (e.g. finding a unique solution in the objective function), we instead focused on the stable region of the model dynamics and used random search method for minimizing the objective function during the estimation. The empirical results show that the fit of the model with multiplicative noise is worse than the model with additive noise. Moreover, we found that ABS can not match the chosen moment conditions regardless of the noise specification. Therefore they fail to explain the empirical regularities of the S&P 500 index.

Concerning the empirical connection between the trading behavior and the price dynamics, we investigated the effects of two trading mechanisms in SSV and compared their empirical performance. To examine to what extent behavioral heterogeneity affects the model dynamics, we tested the two types of trading rules using the formal model comparison. We found that the herding mechanism provides a better fit to the volatility of financial data than does the wealth. However, the null hypothesis of an equal fit of the two models cannot be rejected at 5% level. The variations of the market fraction between heterogeneous investors act as a main driving force behind the non-linear dynamics of SSV. This result shows that the model dynamics of SSV is moderately robust to different specifications for analyzing the trading mechanism.

Appendix A: Stochastic interpretation of switching rule in ABS

In this section, the model dynamics of ABS are extended allowing for a finite number of investors. To examine the finite size effects on the market price volatility, we propose a stochastic guideline to the discrete choice theory; we refine the approximation of the model to the group behavior from the stochastic process.

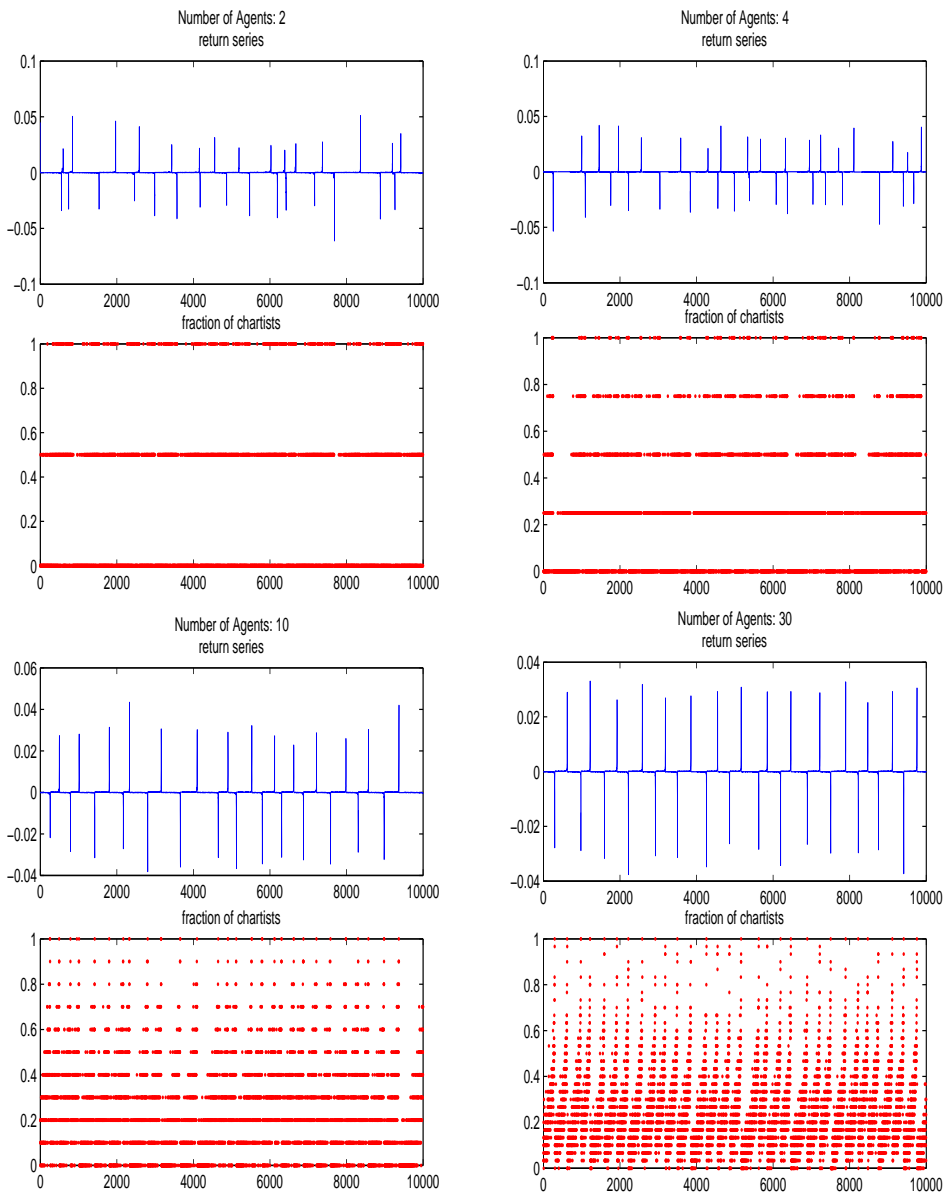


Figure 5.7: Deterministic model dynamics with a finite number of agents

We denote by N_c the stochastic realization of the fractions of chartists from the finite-size system:

$$N_c \sim B(N, n_{2,t}) \quad (5.24)$$

where the variable N denotes the total number of agents used in simulations. $B(\cdot, \cdot)$ denotes the binomial distribution specified by the parameters N and $n_{2,t}$. Figure 5.7 illustrates the finite size effects of ABS on the price dynamics. Note here that the model dynamics with the finite number of agents are contrasted with the infinite benchmark in Equation (5.9); see also Figure 5.1.

It can be seen that irregular jumps in returns are influenced by the misperception of investors: e.g. the random realization of being a chartist in the finite size system. In Figure 5.7 we observe the variation in the return dynamics when a small number of agents is employed in this artificial economy; i.e., $N = 2, 4, 10$. The support of a probability distribution is clearly discrete in this case. For $N = 2$, we can place its probability into three points (0.0, 0.5, and 1.0). The case of $N = 10$ allows for the support of the probability: a set of $\{0.0, 0.1, \dots, 0.9, 1.0\}$. If the number of agents used in simulations increases up to $N = 30$, the price movements follow the periodic limit cycle where the infinite interacting agents are assumed in the model. Overall the misperception of investors is likely to occur when we keep a small number of agents in the system.

B: Selection matrix

The entries of a selection matrix C_W used in the simulated test distribution consist of zeros and unities; see appendix C. For example, we can consider a simple 3×3 positive definite matrix as follows:

$$M = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{12} & e_{22} & e_{23} \\ e_{13} & e_{23} & e_{33} \end{bmatrix}.$$

The matrix M has six unique elements. Then the elements of the matrix has the following relationship (see also the matrix notation of $\text{vec}(W) = C_W \cdot \widehat{\xi}$ in appendix C):

$$\begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{12} \\ e_{22} \\ e_{23} \\ e_{13} \\ e_{23} \\ e_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{22} \\ e_{23} \\ e_{33} \end{bmatrix}.$$

We can easily apply the same procedure to the 10×10 weight matrix in the SMM. The weight matrix used in this study has 55 unique elements. We leave it as a trivial exercise for the reader.

C: Technical note on the model comparison method with a random weight matrix

This section summarizes the matrix notations used in the test distribution of the formal test. The starting point of HMT is that the model is misspecified where a random weight matrix cannot be optimal. Since the use of the weight matrix can influence the asymptotic variance matrix of the method of moments, it is necessary to take into account the correlation between the moment conditions and the consistent weight matrix. That is the main motivation of the HMT's supplementary paper which specifies additional matrix formula for simulating the test distribution.

In this study, we consider a symmetric covariance matrix as the weight matrix (or random weight matrix). The moment estimates and the unique elements of a random weight matrix are asymptotically normally distributed under standard assumptions in HMT-sup¹¹:

$$\sqrt{T} \left((m_T - \widehat{m}(\theta))', (\xi_T - \widehat{\xi})' \right)' \xrightarrow{d} N(0, \widehat{\Sigma}_0) \quad (5.25)$$

where ξ is the unique elements of W . The unique elements of the weight matrix refer to the sampling variability in the moment estimates. For example, the variation in the first moment can be explained by the second moment, while the variation in the variance is represented by the fourth moment (kurtosis). Remember that we can estimate the variation in the unique elements using the double block bootstrap method. The intermediate steps will be discussed in the next section.

¹¹In this study, the terminology of "random" implies that we use a full weight matrix without imposing any restriction on its elements. If this matrix is estimated from the data, we can regard it as a data-dependent matrix.

In our study we set the weight matrix to $W = \widehat{\Omega}_{BB}^{-1}$. To proceed to the next step of matrix computation, we calculate the selection matrix C_W from its simple multiplication of $\text{vec}(W) = C_W \widehat{\xi}$. The matrix C_W has dimension $n_m^2 \times d$ where d denotes the number of unique elements of W . As it is discussed above, we use the block bootstrap method to consistently estimate the variance-covariance matrix $\widehat{\Sigma}_0$ of which dimension is $(n_m + d) \times (n_m + d)$. It is assumed here that the matrices D , F , M are non-singular in neighborhoods of θ :

$$\begin{aligned} D^I &= \left(W \ E_I \otimes (\widehat{m}_T - m^I(\theta^I))' C_W \right) \\ F^I &= \frac{\partial m^I(\theta^I)'}{\partial \theta^I} W \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} - M^I \\ M^I &= (E_I \otimes (\widehat{m}_T - m^I(\theta^I))' W) \frac{\partial}{\partial \theta^{I'}} \text{vec} \left(\frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} \right), \quad I = A, B \end{aligned} \quad (5.26)$$

E_I is the identity matrix of which dimension is $n_\theta^I \times n_\theta^I$. Note here that the dimensions of the matrices $\frac{\partial m^I(\theta^I)'}{\partial \theta^{I'}}$ and $\frac{\partial}{\partial \theta^{I'}} \text{vec} \left(\frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} \right)$ are $(n_m + d) \times n_\theta^I$ and $(n_m + d) \cdot n_\theta^I \times n_\theta^I$. The dimension of F^I and M^I are n_θ^I by n_θ^I .

The theorem S.4 of HMT-sup states that the QLR test distribution converges in distribution to $z' \widehat{\Sigma}_0^{1/2} (U^B - U^A) \widehat{\Sigma}_0^{1/2} z$, where $z \sim N(0, E_{n_m+d})$. The $n_\theta^I + d$ by $n_\theta^I + d$ matrix U^I is defined as $U^I = U_1^I - U_2^I - U_3^I - U_4^I$ with $I = A, B$:¹²

$$\begin{aligned} U_1^I &= D^{I'} \frac{\partial m^I(\theta^I)'}{\partial \theta^{I'}} (F^{I'})^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} W \frac{\partial m^I(\theta^I)}{\partial \theta^{I'}} (F^I)^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} D^I \\ U_2^I &= \left(W \ 0 \right) \frac{\partial m^I(\theta^I)'}{\partial \theta^{I'}} (F^{I'})^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} D^I + D^{I'} \frac{\partial m^I(\theta^I)'}{\partial \theta^{I'}} (F^{I'})^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} \left(W \ 0 \right) \\ U_3^I &= D^{I'} \frac{\partial m^I(\theta^I)'}{\partial \theta^{I'}} (F^{I'})^{-1} (M^{I'} + M^I) (F^I)^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} D^I \\ U_4^I &= \left(0 \ E_I \otimes (\widehat{m}_T - m^I(\theta^I))' C_W \right)' \frac{\partial m^I(\theta^I)'}{\partial \theta^{I'}} (F^{I'})^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} D^I \\ &\quad + D^{I'} \frac{\partial m^I(\theta^I)'}{\partial \theta^{I'}} (F^{I'})^{-1} \frac{\partial m^I(\theta^I)'}{\partial \theta^I} \left(0 \ E_I \otimes (\widehat{m}_T - m^I(\theta^I))' C_W \right) \end{aligned}$$

When two models are overlapping, we can use a sequential testing procedure following Vuong (1989). The first step of the model comparison is to test the null hypothesis that the fit of the models to the empirical moments is equivalent. If the null hypothesis (H_0) is rejected, then the second step is to compute the asymptotic distribution of the QLR statistic (see theorem S.5 in HMT-sup). Since the H_0 cannot be rejected in the first step, we do not proceed to the second step in our study.

¹²The matrix notation $(W \ 0)$ means that the $n_\theta^I \times d$ zero matrix is stacked into the $n_\theta^I \times n_\theta^I$ weight matrix. The dimension of $(W \ 0)$ is $n_\theta^I \times (n_\theta^I + d)$.

D: Double moving block bootstrap method

This section reviews the moving bootstrap method, which is used in this chapter to estimate the covariance matrix; also see Lahiri (2003) for the details. In order to estimate the variations of the moment conditions and unique elements of the weight matrix, we also consider the double block bootstrap method.

We assume that random variables of X_1, X_2, \dots, X_N are weakly dependent and stationary. We denote by χ_n and l the sample and an integer for the sub-sample; i.e. $\chi_n = \{X_1, \dots, X_n\}$, $l \equiv l_n \in [1, n]$. Then we can create the sub-blocked sample of $B_i = (X_i, \dots, X_{i+l-1})$ with the block of length l for X_i , $1 \leq i \leq N$. The intermediate steps are taken as follows:

C1: Generate a bootstrap sample of size T by sampling the data using a fixed block length with replacement. Denote by B^* the resampled data; i.e. $B^* = \{B_1^*, \dots, B_k^*\}$, where the elements in B_i^* consist of $(X_{(i-1)l+1}^*, \dots, X_{il}^*)$, $i = 1, \dots, k$.

C2: Reshuffle B^* again using random interval point and generate a double bootstrapped sample of size T . Denote by B^{**} the double resampled data; i.e. $B^{**} = \{B_1^{**}, \dots, B_k^{**}\}$, where the elements in B_i^{**} are $(X_{(i-1)l+1}^{**}, \dots, X_{il}^{**})$, $i = 1, \dots, k$.

C3: Compute the variance-covariance matrix of B^{**} ; i.e. $\widehat{\Sigma}_0$. This matrix will be inserted into the simulated test distribution for the model comparison (see Equation (5.25) in appendix C).

E: Simulated QLR distribution for model comparison

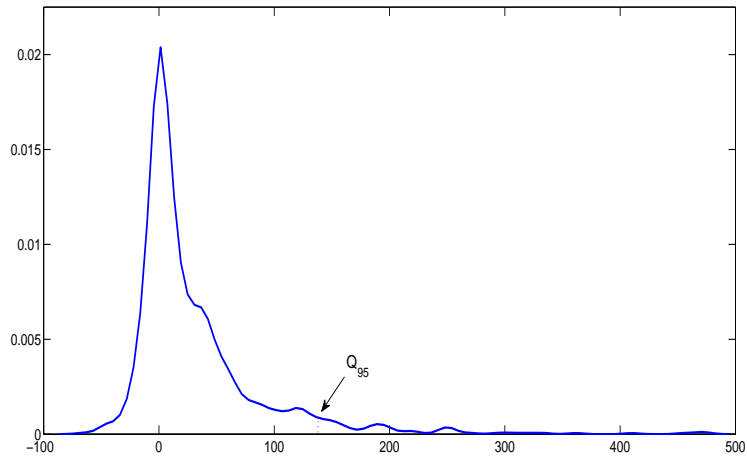


Figure 5.8: QLR distribution of SSV

Social Interaction Effects in Financial Data

In this chapter we estimate a stochastic agent based model (ABM) of an artificial financial market with social interactions among market participants. Assuming that fundamentals are described by a Brownian motion, first, we study the dynamic properties of a simple interacting agent model; i.e. the connection between news innovation and social interactions. Second, we use simulation based inference to identify the effects of the behavior of noise traders on price movements by matching empirical moments of returns (as closely as possible). The results show that the mimetic contagion of investors is an important factor to explain historical prices in five major FX markets. As a result, historical volatility of returns under scrutiny can be robustly decomposed into news innovation (45-55%) and social interaction effects (45-55%).

6.1 Introduction

Financial markets are populated by heterogeneous investors. Their motivations and preferences for investments give rise to complex dynamics in asset markets. However, despite the diversity of expectations and financial objectives, most investors can be categorized as either chartists or fundamentalists.¹ Motivated by an empirical evidence of heterogeneous expectations and trading strategies, a large number of asset pricing models have been proposed allowing for heterogeneity of traders, boundedly rational expectation formation as well as aspects of social interactions; reviews of this literature include Hommes (2006), Lux (2009), Chen et al. (2009) among others.

To name only a few, the adaptive belief system describes an interactive process between different types of traders, who seek for profitable investment strategies; see Brock and Hommes (1997), Gaunersdorfer (2000), Hommes (2001), Chiarella and He (2002), Chiarella et al. (2006), Gaunersdorfer and Hommes (2007). In the model of De Grauwe et al. (1993), a chaotic market force drives price movements in the foreign exchange market; see also De Grauwe and Grimaldi (2006). Their framework for analyzing the behavior of fundamentalists and chartists closely resembles an earlier work of Frankel and Froot (1990). As regards a stochastic approach to

¹The distinction between chartists and fundamental traders has a long history. In particular, early contributions include Baumol (1957), Zeeman (1974), Day and Huang (1990).

group behavior, Kirman (1991, 1993) proposed a simple switching mechanism that mimics the behavior of ants. Lux (1995, 1997) generalized the switching dynamics of the herd behavior implied by a jump *Markov* process. In another model, Westerhoff (2009) as well as Franke and Westerhoff (2009) examined the effect of heterogeneous demand shocks on the economy and explored the dynamics of return volatility according to the structural switching mechanism.

It is assumed in this chapter that behavioral heterogeneity affects price movements in the financial markets. As a possible explanation for heterogeneity, we propose a stochastic model of social interactions and examine, by empirically validating the model, what effect such a non-rational group behavior will have on the market; see Föllmer et al. (2005), Horst and Rothe (2008) among others for a probabilistic approach on the herd behavior in financial markets. Note here that returns would be considered results of an opinion formation process among investors. Then the model and framework in this experimental study show that the behavioral heterogeneity is one of the main driving forces behind price movements. This has some similarities with approaches of other interaction-based models, but a key difference in our approach is our recognition that investment decisions and the communications among agents are endogenously driven by social interactions.

It is shown in Alfarano and Lux (2007) that a simple model of social interactions between optimistic and pessimistic investors can reproduce several salient features about financial time series: volatility clustering, long range dependence and leptokurtic distribution. One difficulty with this approach is, however, that the price dynamics induced by a complex *Markov chain* of the social interaction do not mimic real data. For example, the simulated price bears little resemblance to real prices due to a bi-stable distribution of the opinion formation process in the model; also see Pollard (2006). Moreover, the dynamics of ABMs sometimes rely on unrealistic noise level; e.g. see Lux (2006), Amilon (2008). Such a disconnection between ABMs and the empirical facts can be explained by the model complexity; the modeling approach to ABMs is intuitively appealing (e.g. bottom-up approach), but closed-form expressions of non-linear or stochastic systems are not readily available for the model estimation. Accordingly, the complexity of ABMs becomes a burden for evaluating their empirical performance.

To circumvent this problem, Winker et al. (2007) propose a model validation method using heuristic optimization techniques, which can deal with nonconvex parameter space for estimating ABMs.² Following their study, we demonstrate how to apply the simulated method of moments estimator to empirically validate a simple interacting agent model. To begin with, we use simulation based inference with the block bootstrap method, which decodes a time-dependent structure in financial data; see also Gilli and Winker (2003). Next, we attempt to estimate the model parameters as a direct result of matching the empirical moments of returns (as closely as possible). For that purpose, we propose several moment conditions as stylized facts of financial data; we use them to analyze the time-dependent structure

²They also found that time series models show a better empirical performance than ABMs.

of data using the block bootstrap method, which can allow for an efficient estimation and statistical inference in the context of model misspecification.

The empirical findings of this market microsimulation model show that a mimetic contagion of investors is an important factor to explain the daily return volatility of various foreign exchange (FX) markets. For instance, it is assumed in our artificial economy that the social interaction effects among investors are reflected in a bullish or bearish market sentiment. Especially, a "fragile" (or neutral) market sentiment coincides with erratic price movements in the historical FX market data. This implies that investors relate news arrivals to the future earning of the asset. With a quick update on market information, the noise trader interacts with his or her peers. As a result, the groups of the noise trader are busy to communicate with each other as long as there is much uncertainty about fundamentals in financial markets. Note here that when the FX rallies on bullish (or bearish) sentiment, assets are likely to be overvalued (or undervalued) against the fundamentals. Overall, price fluctuations observed in financial markets can be a direct result of social interactions among market participants.

The remainder of this chapter proceeds as follows. Section 2 reviews a simple model of the opinion dynamics and discusses the properties of an artificial financial market with social interactions. To set the stage for our subsequent empirical analysis about the trading behavior from various FX markets, section 3 discusses the simulated method of moments, the block bootstrap method, and identification problems in detecting group behavior of the model. Section 4 presents main results of a model validation and empirical estimates obtained from the data. Finally, section 5 concludes. Appendix collects all relevant details.

6.2 The model framework

6.2.1 The behavior of noise traders

In this section, we discuss the basic properties of group behavior in a simple interacting model. To begin, the sentiment of the market is represented as being in an opinion formation process among N agents; i.e. the redefinition of a binary opinion variable, say $+1$ (optimistic) and -1 (pessimistic) stands for the opinion of each agent. The variables $N_{O,t}$ and $N_{P,t}$ denote the number of optimistic and pessimistic agents at date t , respectively. Next, the state of the system is defined as the market sentiment (opinion index):

$$x_t = \frac{N_{O,t} - N_{P,t}}{N} \quad (6.1)$$

Each agent revisits and possibly changes his or her opinion according to a Poisson process with intensity ν . The probability for the opposite opinion is assumed to be proportional to the current fraction of agents maintaining the opposite view of the

market. Agents eventually reconsider their opinion and evaluate the current market sentiment, which is reflected in the opinion index. Then a pessimistic agent decides to join the optimistic group with a conditional probability $\omega(N_O + 1, N_P - 1 | \cdot)$, whereas the reverse move is obtained with $\omega(N_O - 1, N_P + 1 | \cdot)$. In a continuous-time framework, the state configuration of the market during an infinitely small time interval Δt is governed by the following transition probabilities:

$$\begin{aligned}\omega(N_O + 1, N_P - 1, t + \Delta t | N_O, N_P, t) &= \nu \Delta t N_P \frac{N_O}{N} \\ \omega(N_O - 1, N_P + 1, t + \Delta t | N_O, N_P, t) &= \nu \Delta t N_O \frac{N_P}{N} \\ \omega(N_O, N_P, t + \Delta t | N_O, N_P, t) &= 1 - 2\nu \Delta t \frac{N_O N_P}{N}\end{aligned}$$

where the parameter ν might now be seen as a combined constant for Poisson clock and the proportionality factor for the strength of social interactions. The state of the system evolves over time as agents switch between groups.

To make the process more explicit, we denote by X and Y the random variables of a binomial distribution:

$$\begin{aligned}X &\sim B(N_O, t, w) \\ Y &\sim B(N_P, t, w)\end{aligned}\tag{6.2}$$

where X and Y are independent replications from the transition probability. Note here that a binomial random variable can be most easily simulated by the sum of N independent Bernoulli random variables. Hence it holds that

$$\begin{aligned}N_{O,\Delta t} &= N_{O,t} - X + Y \\ N_{P,\Delta t} &= N - N_{O,\Delta t}\end{aligned}\tag{6.3}$$

This implies that a binary choice at individual level explains macroscopic behavior in this artificial economy. To prevent the simple *Markov chain* process from being absorbed at the boundaries of $x = \pm 1$, the following reflecting rules are added:

$$\begin{aligned}\text{If } N_P(t) = 0, \text{ then } P\{N_P(t + \Delta t) = 1\} &= 1 \\ \text{If } N_O(t) = 0, \text{ then } P\{N_O(t + \Delta t) = 1\} &= 1\end{aligned}\tag{6.4}$$

This rule makes the extreme states of the opinion index ($x = \pm 1$) transient, but leads to a distribution that has its peaks at the boundaries (U-shaped distribution). Accordingly, the model has a unique equilibrium in the path of opinion dynamics; see Alfarano and Lux (2007).

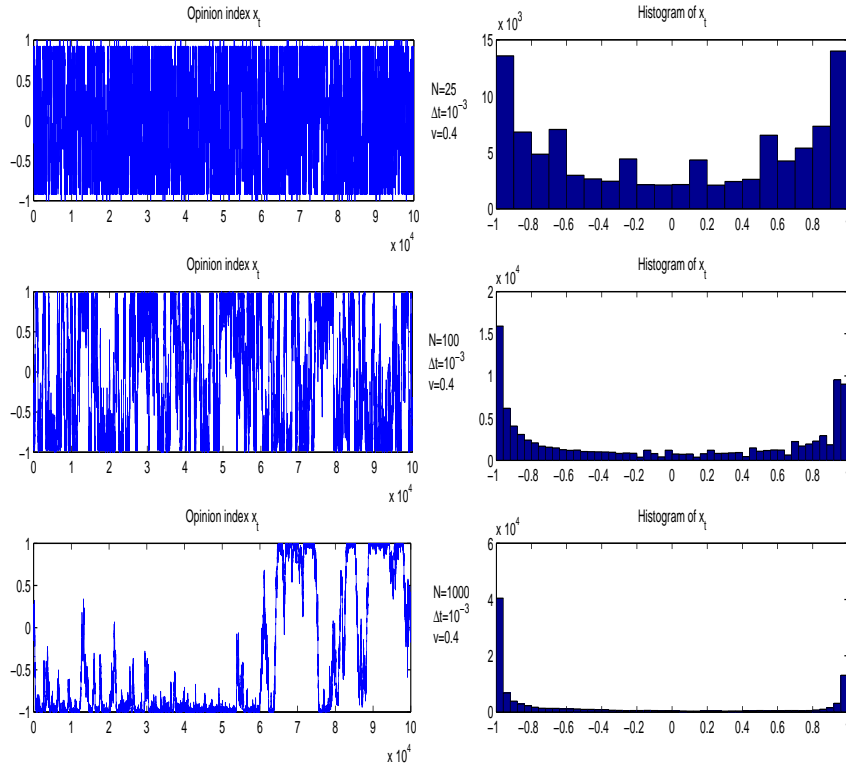


Figure 6.1: The opinion index and its frequency histogram

Note: The upper, middle and lower panels are simulated using a small ($N=25$), large ($N=100$), and extremely large ($N=1,000$) system size.

The model takes as input binary decision processes that specify the development of group dynamics with respect to the system size used in the economy. Concerning the model estimation, we use a large number of agents N to maintain an appropriate level of social interactions. Indeed, Alfarano and Lux (2007) proved the existence of the unique equilibrium distribution of an opinion formation process, of which functional form is U-shaped:

$$P_e(x) = \frac{1}{L} \frac{1}{1-x^2} \quad (6.5)$$

where L is a normalization constant:

$$L = \ln \frac{2-\delta}{\delta}, \quad \delta = \frac{1}{N} \quad (6.6)$$

Therefore the dynamics of opinion formation in the equilibrium distribution are dependent on the current opinion index x_t and the number of agents N . The relative frequency of the opinion index is influenced by the system size of this artificial

economy. Indeed, the large system size ($N \rightarrow \infty$) will reduce the observed frequency of the intermediate state in the opinion index ($-0.5 < x_t < 0.5$), but puts more emphasis on the two bi-stable states ($x_t = \pm 1$). This relation between the system size and the opinion index is attributed to the fact that the change of the normalization constant in Equation (6.5) affects the equilibrium distribution.³ This implies that the opinion formation process is dependent on the number of agents; see Figure 6.1. In general, we fix N to a large value for the model estimation, and investigate the effects of the system size on the parameter estimates. This will be discussed later.

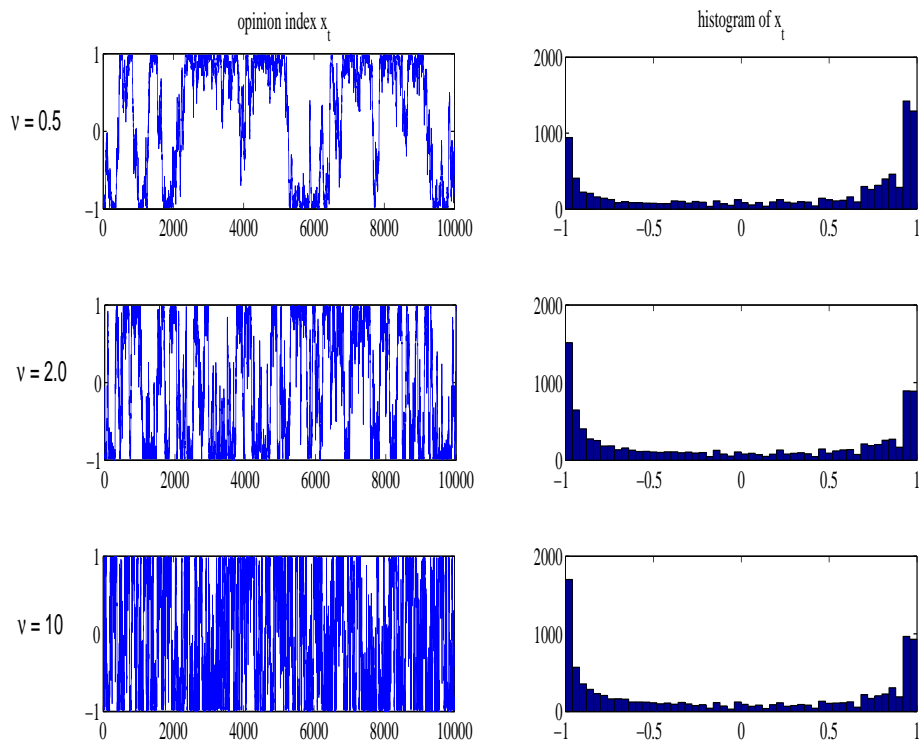


Figure 6.2: The behavior of switching process in a noise trader model

Note: The upper, middle and lower panels are simulated using the parameter $\nu = 0.5, 2.0, 10$, respectively. The other parameters are fixed: $\beta = 1.0$, $N_F T_F = 10$, $N_C T_C = 1$, $100 \cdot \delta_0 = 0.4$, $\Delta t = 0.01$, $N = 100$.

It is noteworthy to remember that the parameter ν , which controls the speed of the group behavior, has no impact on the qualitative dispersion of the equilibrium distribution; see Figure 6.2. However, this parameter value is associated with the

³The variation in the equilibrium distribution is controlled by the logarithmic transformation of the system size. For instance, $L = \ln \frac{1.99}{0.01} \simeq 5.29$ for $N = 100$, while we arrive at the normalization constant L of 7.60 for $N = 1000$.

velocity of the switching process at micro-time step. Indeed, the increase in the parameter ν affects the switching process in a highly dynamic way and gives rise to more extreme opinions observed in the market.

6.2.2 The artificial market and benchmark simulation

Now we relate a composition of the population of the investors to the log-returns of the market price; see Alfarano et al. (2008). Note here that the fundamentalists adopt a mean reversion behavior when they respond to excess demand. On the contrary, noise traders exploit investment opportunities from the market sentiment. Assuming that a market maker adjusts the price according to the observed excess demand, we can write a stochastic differential equation implied by the news innovations and the opinion dynamics.

$$\frac{1}{p_t} \frac{dp_t}{dt} = \beta \cdot [N_F T_F \cdot \ln\left(\frac{p_F}{p_t}\right) + N_C T_C \cdot x_t] \quad (6.7)$$

where the parameter β measures the speed of price adjustments. The parameters $N_F T_F$ and $N_C T_C$ denote the transaction volumes of the fundamentalists and the noise traders, respectively. For instantaneous market clearing ($\frac{dp_t}{dt} = 0$), we may write the price equation as:

$$\ln p_t = \ln p_{F,t} + \frac{N_C T_C}{N_F T_F} \cdot x_t \quad (6.8)$$

Because the purpose of the microsimulation model is to explain the empirical observation of historical financial data, we assume that fundamentals are described by a Brownian motion. As a source of randomness in the model, the market's response to the news arrival influences the price dynamics.

$$d \ln p_{F,t+\Delta t} = \delta_0 \cdot \varepsilon_t dt \quad (6.9)$$

where ε_t is a standard Gaussian random variable at time t , and the constant term δ_0 denotes a measure of the news arrival in the market. The news innovation in Equation (6.9) with transition rates for the opinion index establishes the connection between the opinion index and the price fluctuations. Then the model predicts that the groups of noise traders are busy to communicate with each other when the market mood is moderately neutral (i.e. $-0.5 < x_t < 0.5$).

Turning to the relation between social interactions and the fraction of optimistic/pessimistic agents, the model predicts that the social interactions will have strong effects on price movements as long as the opinion of two groups is starkly

divided over the market regime. If the half of agents is optimistic and the other half is pessimistic (i.e. $N_O = N_P$), the transition rate, which describes the changes in the socio-economic configurations, arrives at the highest point; i.e. the numerical value will be $\nu\Delta t\frac{N}{4}$. Accordingly, before they sell or buy assets, investors make an effort to collect information about fundamentals and are likely to cope with peer pressure, which can influence their decision making. On the other hand, the transition rate of the boundary points (i.e. $N_O = N_P = 0$) arrives at the minimal value of $\nu\Delta t\frac{2N-1}{4N}$ instead of no movements. In sum, this observation summarizes the link between the transition rates of the group behavior and the price dynamics.

Table 6.1: Parameter values for simulation

	ν	N_FT_F	N_CT_C	$100 \cdot \delta_0$	β	Δt	$\log p_{F,0}$	N
value	0.5	10	1.0	0.4	1.0	0.01	1.0	100

However, the price movements with the opinion index can not readily be expressed in a closed form way; e.g. Alfarano et al. (2007) derived analytic moments of the variant of this simple interacting model, but in our case this mathematical derivation is much more complicated due to the combination of news innovation and the group behavior. Instead of relying on the analytic solution of the model, hence, we simulate the price path p_t by discretizing Equation (6.7) for estimation. We assume that time is discretely defined over the trading days ($\Delta t = 0.01$). Now the solution of the price adjustment process can be approximated using an implicit Euler scheme.⁴

In this simulation, we use 100 noise traders and allow for social interactions at each micro-time step. For the sake of simplicity, the adjustment speed parameter β is set to unity. Since the adjustment speed parameter is fixed due to the collinearity (i.e. the price equation multiplies this parameter and the other trading volume parameters, so these parameters β , N_FT_F , and N_CT_C cannot be identified simultaneously), we attempt to estimate the other parameters; i.e., N , ν , N_FT_F , N_CT_C , δ_0 . As the opinion dynamics of the complex *Markov Chain* are influenced by the parameters ν and N , however, the model generally faces identification problems in the estimation procedure; the mapping between these parameters and the objective function is highly non-linear, and the change of these parameter values can induce different trajectory of the price dynamics. As regards the estimation strategy, for example, we fix the parameter N (i.e. the system size), which makes the path of the price dynamics stable in the model. In particular, we examine the importance of the system size for evaluating the fit of the model and its influence on parameter estimation. The details of identification issues in this model will be discussed later.

⁴If a small step size (i.e. $\Delta t = 0.01$) is used for approximating the solution of the price dynamics, then the computation will be very expensive. For instance, in order to generate 10,000 price series, $10,000 * \frac{1}{0.01}$ times of computations have to be executed. Note here that we translate a main algorithm for the switching process into a powerful C language to enhance execution speeds.

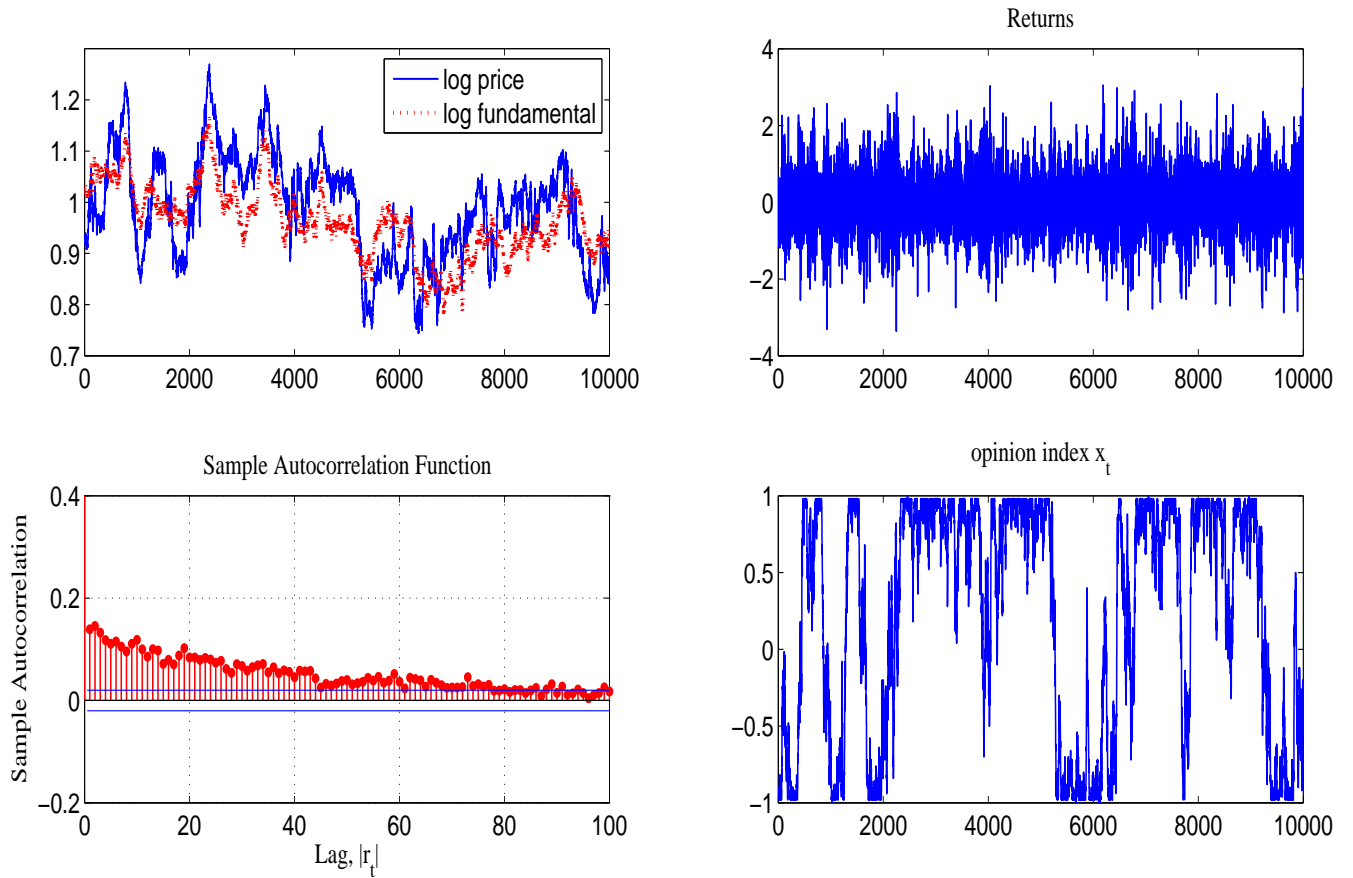


Figure 6.3: Noise trader model with social interactions

Note here that we simulate the model by using the parameters documented in Table 6.1. The model synchronizes a dominance of the optimistic or pessimistic investors with the over- or under-valuation of assets (see Figure 6.3). Since the social interaction effects are strong for the transition periods between the extreme market regimes, it is shown that the social interaction proxied by the opinion index has a direct influence on the market turbulence: the stronger social interactions will lead to the higher volatility in financial markets. The autocorrelation pattern of the absolute returns appears highly persistent.

6.3 Simulation based inference

In this section, first, we explain the simulated method of moments (SMM) for estimating the group behavior. Second, we present several sets of moment conditions

used in the SMM estimation. Finally, we discuss identification problems in the model by investigating three-dimensional parameter space.

6.3.1 The simulated method of moments estimation

The SMM estimation is the counterpart of the generalized method of moments when theoretical moments are approximated by simulations.⁵ Returns of the simulated data series are defined as:

$$r_{t+1} = 100 \times \ln\left(\frac{p_{t+1}}{p_t}\right) \quad (6.10)$$

We consider the price equation that is generated by the exogenous opinion index (x_t) and the other parameters of the trading behavior; $p_t = f(p_{t-1}, x_t, \varepsilon_t; \theta)$. The variable ε_t denotes idiosyncratic shocks of a stochastic process at date t . Note here that the parameter vector θ has a dimension of $m \times 1$. Since the stochastic process for p_t results from a combination of the opinion dynamics spelled in Equation (6.1) to (6.4) and the stochastic differential equation for the price adjustment (Equations (6.7) and (6.9)), a conditional likelihood function of the model $f(p_t | z_t; \theta)$ does not have a simple analytic solution with $z_t = [p_{t-1}, \dots, p_0]'$. Alternatively, we can approximate the model solution based on simulations. The distance of the model-generated moments (f_t) from the empirical moments (f_t^*) is defined as:

$$D(\theta) = \frac{1}{T} \sum_{t=1}^T f_t^* - \frac{1}{ST} \sum_{t=1}^{ST} f_t \quad (6.11)$$

where S denotes the simulation size, and its high value asymptotically reduces the simulation error. The simulations can be made to approximate the population moments of a structural model by performing a Monte Carlo integration as determined by $\frac{1}{ST} \sum_{t=1}^{ST} f_t$; see also Michaelides and Ng (2000) among others. To obtain an approximation of moments conditions with desired precision, we fix the simulation size to 100. The first 500 simulation data are discarded to exclude an initial transient period to ensure that p_0 has been drawn from a stationary distribution for p_t (Lee and Ingram (1991)). In practice, we can simulate S independent sequences of a length T rather than a single sequence of ST , because this can reduce simulation errors of a highly persistent data; see Jonsson and Klein (1996).

If the model is well specified to generate the data for some $\theta \in R^m$, the resulting estimator becomes consistent (Duffie and Singleton (1993)).

$$\text{plim}_{T,S \rightarrow \infty} D(\theta) = 0 \quad \Bigg|_{\theta = \theta_0} \quad (6.12)$$

⁵See Carrasco and Florens (2002) for a detailed survey on the simulated method of moments.

where D is a $(n \times 1)$ vector of the moment conditions. We minimize the objective function. At its minimum the model can match the empirical moments as close as possible. Note here that the objective function is composed of the difference between the simulated and actual data with a positive definite weight matrix W .

$$\Psi(\theta) = \min_{\hat{\theta}} \hat{D}' W \hat{D} \quad (6.13)$$

As long as a system of estimating moment conditions over-determines the unknown parameters being estimated ($n > m$), this can provide a basis for an over-identifying restrictions test. More generally, if an estimated value for $\hat{\Psi}$ is greater than a critical value given by the χ^2 distribution, then we will reject the null hypothesis that the model is correctly specified. A threshold value for the test is drawn from the right-skewed distribution, in which degrees of freedom are equal to the difference along the dimensions of D and θ ; i.e., $n - m$.

In the present study, we use the block bootstrap resampling method to compute the weight matrix following Winker et al. (2007). The empirical data are resampled 1,000 times from which we estimate their variance-covariance matrix and compute the test distribution of the model misspecification. Standard hypothesis tests and statistical inference will be discussed using the block bootstrap method later.

6.3.2 The moment conditions and the block bootstrap method

In the current study, we utilize the moment conditions covering the main features of financial data; e.g. see chapter 4 of Taylor (2005) regarding stylized facts of stock returns and Winker and Jeleskovic (2006, 2007) for exchange rate returns. Our moment statistics are summarized in Table 6.2; see also appendix A.⁶ The DM/US data set covers a period from Jan.02.1973 to Dec.11.2009 at a daily frequency. The sample size is 9,000; see Figure 6.4.

The mean and the second moment (m_1, m_2) measure the shape of the return distribution. The sample autocorrelation (ACF) of raw returns at lag 1 (m_3) provides some indication of almost no correlation between returns for different days. The ACF of the absolute returns at lag 1, 5, 10, and 20 ($m_4 - m_7$) shows a highly persistent nature of volatility.

The selected ACFs of $|r_t|$ at lags provide partial information about the presence of serial correlation, but the parameters of auxiliary models capture global time-dependence properties of financial time series. For example, the GARCH (1,1) model describes a presence of a short-range dependence in the data. Note here

⁶Following Winker and Jeleskovic (2006, 2007), we attempted to consider temporal aggregation of the FX data (e.g. from daily to weakly frequency). However, the inclusion of such moment conditions may distort information about time-dependent structures in the data due to the shrinking number of observations.

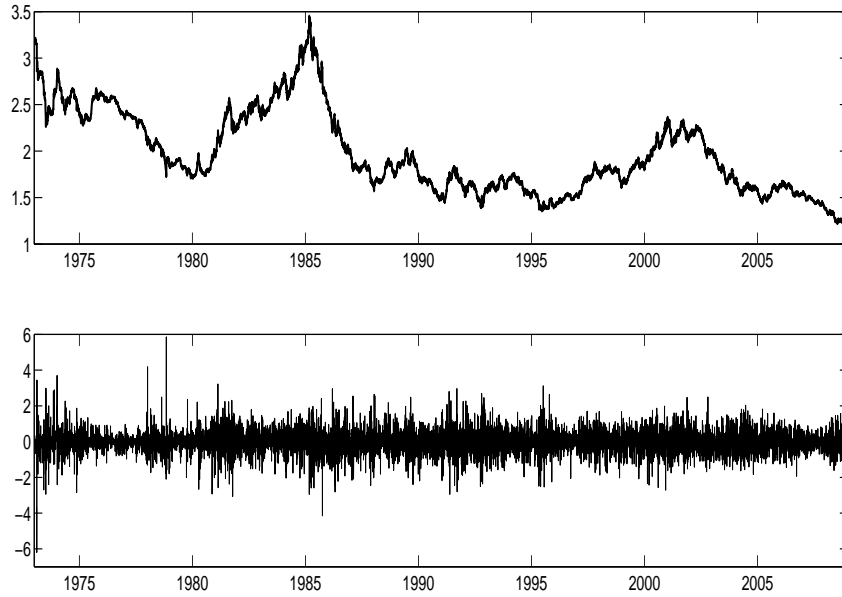


Figure 6.4: Prices (upper) and returns (lower) of historical DM/US

that the moment $m_{ID,1}$ (α_1) measures the extent to which the current volatility feeds through into the next day's volatility. The coefficient β_1 ($m_{ID,2}$) captures the persistence of the shock over time in the return volatility. Finally, the tail index measures extreme events of the return distribution ($m_{ID,3}$). The statistic entering the objective function is the mean value of the Hill estimator over the 5-10% upper quantiles of the right tail. A threshold is selected from an estimated plot where a shape parameter α is fairly stable. If a low threshold value is included, this will improve the accuracy of the obtained estimates, but these become biased; see Figure 6.5.

With a focus on these moment conditions, we can compute their covariance estimates using the bootstrap method, of which inverse form can serve as the weight matrix in SMM. To provide an explanation of intermediate computational steps, we introduce notations; we denote by $f^b \in R^b$ a vector for the sample moments b . Note here that B is the number of block bootstrap resampling, \bar{f} is the mean value for the bootstrapped moments. The weight matrix in the objective function is obtained by taking the inverse of the variance-covariance from 1000 block bootstrap samples:

$$\hat{\Omega}_{BB} \equiv Var(m^b) = \frac{1}{B} \sum_{b=1}^B (f^b - \bar{f})(f^b - \bar{f})' \quad (6.14)$$

Table 6.2: Statistics of empirical moments and block bootstrap samples: DM/US

Label	Statistic	Empirical	Bootstrap
m_1	$E[r_t]$	-0.0089	-0.0091
m_2	$\text{Var}[r_t]$	0.4241	0.4234
m_3	$\text{corr}(r_t, r_{t+1})$	0.0287	0.0271
m_4	$\text{corr}(r_t , r_{t+1})$	0.1386	0.1385
m_5	$\text{corr}(r_t , r_{t+5})$	0.1453	0.1430
m_6	$\text{corr}(r_t , r_{t+10})$	0.1253	0.1227
m_7	$\text{corr}(r_t , r_{t+20})$	0.0951	0.0891
$m_{ID,1}$	α_1 (GARCH(1,1))	0.0799	0.0730
$m_{ID,2}$	β_1 (GARCH(1,1))	0.8900	0.8995
$m_{ID,3}$	α (Right tail, 5 - 10 %)	3.3781	3.4049
Test criterion		5%: 18.96	1%: 24.14

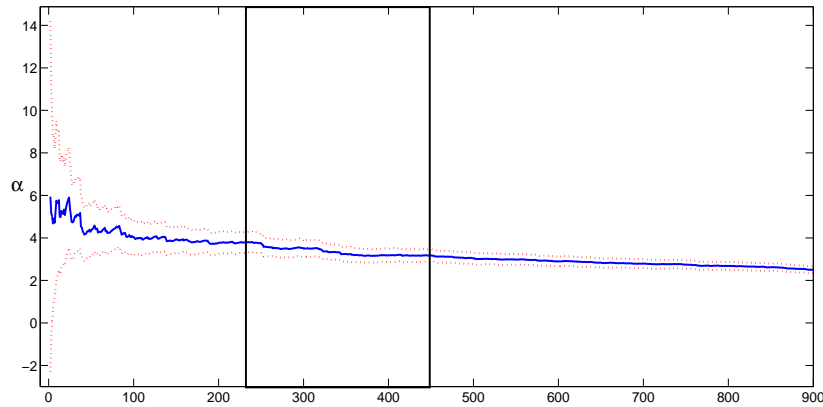


Figure 6.5: Hill plot for daily returns

Accordingly, we can incorporate the weight matrix of $\widehat{W} = \widehat{\Omega}_{BB}^{-1}$ into the objective function in Equation (6.13).

6.3.3 The algorithm and the hypothesis test

Before minimizing the objective function in Equation (6.13), we have to simulate price series from the model, and compute the chosen moment conditions. Below is an example of simulation based inference with a target parameter setting ($N_F T_F$, $N_C T_C$). We implement the following algorithm (1) to estimate the moments over a grid of parameters ⁷:

⁷Our algorithm follows the approach by Gilli and Winker (2003).

Algorithm (1): generating stochastic price process and estimating moments

Partition the grid point of parameters as $[N_F T_F, N_C T_C] \equiv [\{N_F T_F^1, \dots, N_F T_F^k\}, \{N_C T_C^1, \dots, N_C T_C^j\}]$, where $N_F T_F^k \in R^p$ and $N_C T_C^j \in R^q$. The price series p_t is given by $p_t^i \in R^r$. Common random numbers are used to make this experiment.

1. for $k = 1 : R^p$ in the parameter value of $N_F T_F$ do
2. for $j = 1 : R^q$ in the parameter value of $N_C T_C$ do
3. for $i = 1 : R^r$ do
4. generate random sequences for price simulation
5. simulate price path
6. end for
7. compute returns and estimate moments
8. end for
9. end for

After obtaining the moments from the simulated data, we use the algorithm (2) to calculate a value for the objective function. We examine the variations of the objective function Ψ by using several starting values.⁸

Algorithm (2): computing objective function

Given the grid points over $N_F T_F^k \in R^p$, $N_C T_C^j \in R^q$

1. for $k = 1 : R^p$ in the parameter value of $N_F T_F$ do
2. for $j = 1 : R^q$ in the parameter value of $N_C T_C$ do
3. obtain a value for the objective function Ψ
4. end for
5. end for

The algorithms (1) and (2) produce a three-dimensional objective function image over the grid of two parameters. Figure 6.6 depicts the response surface of Ψ over the parameter pairs, (N, δ_0) and $(N_F T_F, N_C T_C)$. The simulated response surface is smooth and convex with respect to the pair $(N_F T_F, N_C T_C)$, but it is non-convex and discontinuous with the pair (N, δ_0) .

Keeping the other parameters fixed for the pair $(N_F T_F, N_C T_C)$, that is $(\nu, N, \Delta t, \beta, 100 \cdot \delta_0) = (1.0, 200, 0.01, 1.0, 0.4)$, we obtain a minimum value for Ψ (around 241) at the point $(N_F T_F, N_C T_C) = (23, 2.1)$. Similar steps are taken to estimate the pair $(N, 100 \cdot \delta_0)$. In this case, the other parameters are fixed; $(N_F T_F, N_C T_C, \Delta t, \beta) =$

⁸In optimization, we implement both the random search and Nelder-Mead simplex methods. If there is no prior information about a parameter range, the random search algorithm seems to be efficient, but it is more time consuming than the simplex. The simplex method is efficient to search for the lowest values of the objective function at a local area.

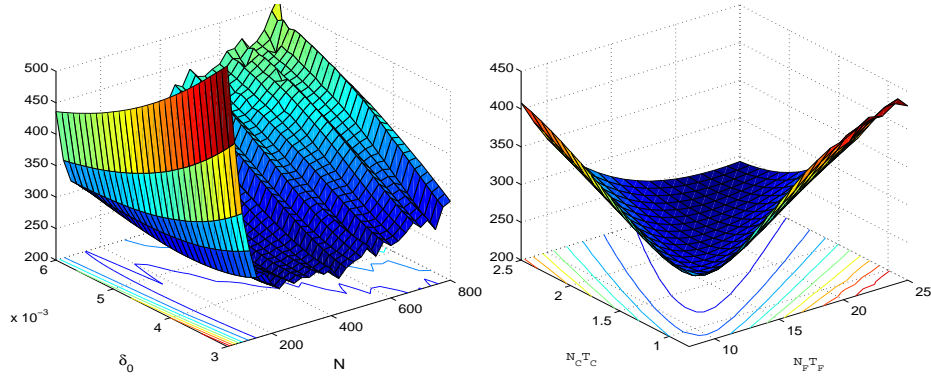


Figure 6.6: Three-dimensional contour plot of objective function

Note: We use two pairs of parameter combinations: left panel for (N, δ_0) and right panel for $(N_F T_F, N_C T_C)$.

$(10, 1.0, 0.01, 1.0)$. Now we find a minimum value of 255 at the point $(N, 100 \cdot \delta_0) = (150, 0.45)$.

To assess the fit of the model to the data, we use a criterion from the block bootstrap method and compare it with the estimated loss function value $\widehat{\Psi}$ of the model. By applying the block bootstrap method, we can obtain the following distribution for the hypothesis testing:

$$J_b \equiv \{(m^b - \bar{m})' \widehat{W} (m^b - \bar{m}) : b = 1, \dots, B\} \quad (6.15)$$

Figure 6.7 illustrates a kernel estimate for the objective function from 1,000 bootstrapped J_b . It is shown that the quantile distribution has a larger dispersion than the χ^2 -distribution with $n - m$ degrees of freedom. For instance, from the DM/US data, we can compute the model selection criteria; i.e., 26.12 at 5% level and 36.52 at 1% level. These values will be used to test whether the model-simulated data are drawn from the true data generating process. If the objective function value exceeds these criteria, the model cannot match the data along certain dimensions of the empirical distribution. In other words, we conclude that the model is misspecified.

However, the econometric analysis cannot offer a reliable evaluation on the fit of the model, because the objective function may exhibit multiple local minima in the parameter space; see the left panel of Figure 6.6. In order to find a global minimum in the multidimensional parameter space, the search process of the optimum is restricted to a local level. The parameter identification problems in this model will be addressed in the next section.

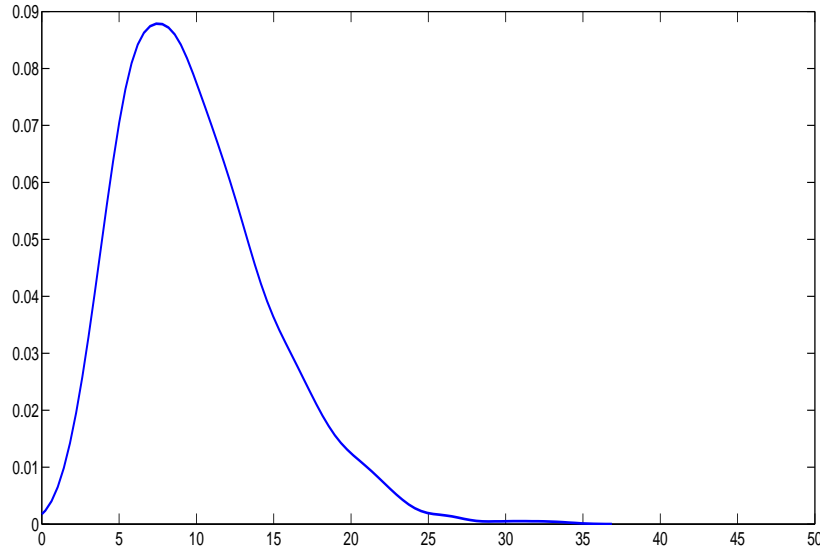


Figure 6.7: Distribution of objective function values

Note. The figure is drawn from 1,000 block bootstrapped samples

6.3.4 Collinearity and parameter identification

The opinion index in the model is governed by the *Markov Chain* process, which is time-homogeneous and geometrically ergodic. Note here that the herd tendency parameter ν determines the threshold level, which controls the number of jumps (birth and death) at each (micro-level) time step. However, we observe the discontinuous response of the opinion index with respect the parameter ν when approximating this switching process. For example, the dynamics of the opinion index can be approximated by a recursive stochastic differential equation (*Langevin*):

$$dx_t = \sigma(x_t) \cdot \eta_t dt, \quad \sigma(x_t) = \sqrt{\frac{\nu \Delta t (2 - \nu \Delta t)}{N} [1 - x_t^2]} \quad (6.16)$$

where η_t is a standard Gaussian random variable at time t , and the time-varying term $\sigma(x_t)$ measures the diffusion process of group behavior.

Moreover, the process of opinion formation between agents is controlled by the jump component x_t . Therefore the simulated response of the objective function generally exhibits discontinuities. This happens that a small change in the parameter ν causes *jumps* in the opinion index, and these jumps, on their turn, cause jumps in the objective function through the differential equation of the price dynamics. In optimization, the Nelder-Mead search method can handle conditions such as discontinuities and multimodalities at the local level if we should cope with the *jumps* in the objective function.

Table 6.3: Grid points in three-dimensional contour plot

label	sub-interval of grid points	stepsize
ν	$0.35 + i * h: 0.35 - 0.65$	$h = 0.01, i = 0, \dots, 30$
$N_F T_F$	$6.5 + i * h: 0.35 - 0.65$	$h = 0.25, i = 0, \dots, 30$
$N_C T_C$	$0.65 + i * h: 0.65 - 1.25$	$h = 0.02, i = 0, \dots, 30$
$100 \cdot \delta_0$	$0.25 + i * h: 0.25 - 0.55$	$h = 0.01, i = 0, \dots, 30$
N	$85 + i * h: 85 - 115$	$h = 1.0, i = 0, \dots, 30$

To check the presence of discontinuity associated with the parameters of this model, we investigate the objective function by drawing its three-dimensional parameter space. In this simulation we set the simulation size to 10. Table 6.3 presents the grid points over the parameters; Figure 6.8 illustrates three-dimensional contour plots over the grid points of possible parameter combinations (i.e., the parameter ν and the other parameters such as $N_F T_F$, $N_C T_C$, δ_0 , N). This confirms that the discontinuity locations are primarily caused by the parameters ν and N .

Similarly, the stochastic group dynamics aggravate the identification problems in the model due to the flatness of the parameter space. For instance, if the jump intensity ($\nu \uparrow$) increases, then the decrease in the trading volume parameter for the noise traders ($N_C T_C \downarrow$) offsets the effects on the price volatility. As a result, the volatility of returns changes a little; the upper right panel of Figure 6.8 shows that the objective function values are similar along the diagonal lines on the graph. For example, the objective function from some combinations of trading volume and herding tendency parameters yields almost the same numerical values: $(N_C T_C, \nu, obj.) = (1.19, 0.4, 37.04), (0.77, 0.58, 38.80)$.

Moreover, an increase of the jump intensity parameter is likely to be proportional to the variation in the trading volume parameter for the fundamentalists. Indeed, the increase in the jump intensity ($\nu \uparrow$) affects the occurrence of the extreme opinions (i.e. near the points $x_t = +1$ or -1) and reduces the observation of the opinion index near the point $x_t = 0$; see Figure 6.2. Therefore this effect can be offset by an increase in the trading volume parameter for the fundamentalists ($N_F T_F \uparrow$). We can simply check a proximity to the multi-collinearity between the parameters of the trading volume and the jump intensity using a 3-D contour plot; see the valleys along the parameters in the upper left panel of Figure 6.8.

Because of this long "valley" in the objective function surface, a gradient-based numerical optimization will stop at a more or less random point in it. Even more, since the objective function is non-convex in the space direction of the parameter ν , it may have multiple local minima. Therefore, the optimization algorithm will stop in one of these local minima of an inferior objective function value. Overall, from Figure 6.8 we can infer that the jumps of the switching mechanism and the relation to the trading volume parameters give rise to multimodality and associated problems for statistical inference. If multimodality does occur, (1) there are several solutions for the parameters of trading volume, news innovation and herding tendency. (2) The starting values in optimization will be a local minimum during the estimation.

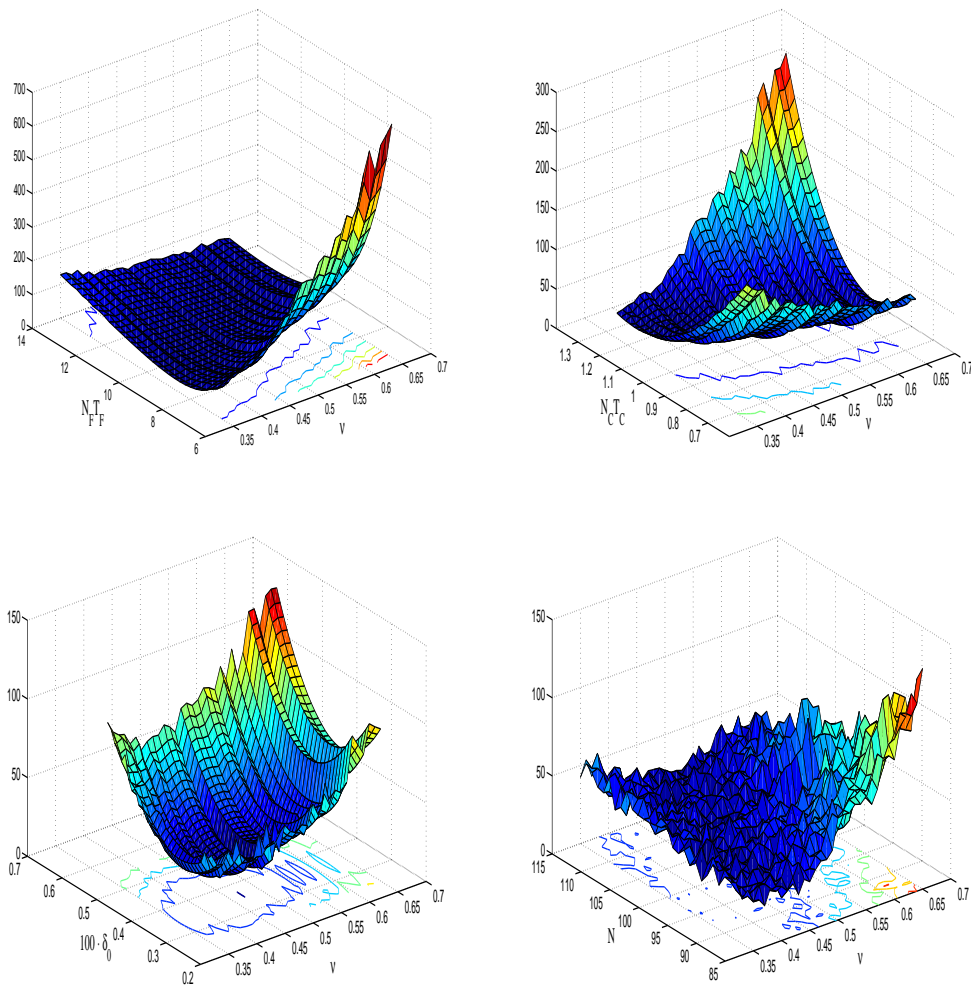


Figure 6.8: 3-D contour plots

Note. Upper left panel plots 3-D surface of the parameters between N_{FF}^{TT} and ν while upper right panel plots N_{CC}^{TT} and ν . Lower left panel plots 3-D surface of parameters between $100 \cdot \delta_0$ and ν while lower right panel plots N and ν .

Table 6.4: Flatness and multimodality I of objective function

Label	sv	estimate	sv	estimate	sv	estimate
ν	0.5	0.543	0.5	0.526	0.4	0.426
$N_F T_F$	5.0	11.201	7.5	8.532	7.5	8.783
$N_C T_C$	0.5	1.033	0.5	0.803	1.25	0.909
$100 \cdot \delta_0$	0.5	0.369	0.5	0.367	0.4	0.395
Ψ		22.20		23.74		18.09

Note: The abbreviation "sv" stands for starting value used in optimization. The system size N is set to 100.

Indeed, it is commonly observed that different minimum values are found in optimization when we change the starting values. It is shown in Table 6.4 that the obtained objective function values does not have a big difference between the parameters ν and $100 \cdot \delta_0$. For example, the model with the parameters of ($\nu = 0.54$, $100 \cdot \delta_0 = 0.37$) and ($\nu = 0.43$, $100 \cdot \delta_0 = 0.40$) has the loss function values for $\Psi = 22.20$ and 18.09, respectively. The proximity to the parameter collinearity is also reported in Table 6.5. This still holds in the relation between the trading volume and herding tendency parameters; i.e. $\Psi = 30.64$ for $(\nu, N_F T_F) = (0.53, 16.6)$ while $\Psi = 31.12$ for $(\nu, N_F T_F) = (0.57, 8.0)$.

Table 6.5: Flatness and multimodalities II of objective function

Label	sv	estimate	sv	estimate	sv	estimate
ν	0.5	0.534	1.5	1.516	0.6	0.567
$N_F T_F$	15.0	16.595	15.0	9.924	6.5	8.002
$N_C T_C$	1.5	1.529	0.5	0.526	0.85	0.727
$100 \cdot \delta_0$	0.5	0.389	0.5	0.488	0.4	0.389
Ψ		30.64		53.02		31.12

Note: The abbreviation "sv" stands for starting value used in optimization. The system size N is set to 100.

Next, we measure the variation in the objective function value with respect to the parameter ν . The other parameters are set: $(N_F T_F, N_C T_C, \delta_0, N) = (10, 1, 0.4, 100)$. From Figure 6.9 we can see that the objective function is poorly behaved with respect to the parameter ν ; see a wiggle in parameter space. The global shape of the objective function is concave, but the function exhibits multiple local minima. If simulation and discretization errors of the function approximation are taken into account, a objective function value ($\Psi = 25.53$) at the point of $\nu = 0.41$ is numerically indistinguishable from the point of $\nu = 0.49$ ($\Psi = 24.42$). Accordingly, the SMM estimation often provides inferior objective function values with its parameter estimates during the minimization of the objective function.

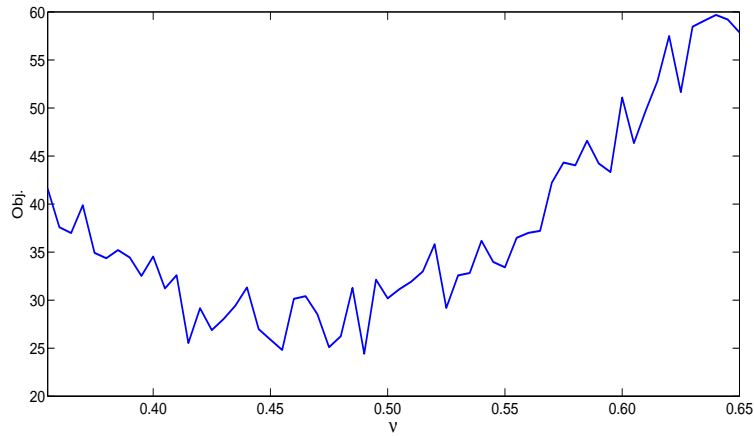


Figure 6.9: The behavior of objective function w.r.t the parameter ν

Note: The system size N is set to 100.

These empirical results show that finding a global minimum in SMM is non-trivial and computationally difficult for this model unless we fix some of the parameters. Accordingly, we can fix the parameter ν in our empirical application and estimate the other parameters in the model. As an alternative solution, we can also avoid this identification problem by setting the trading volume parameters to the values associated with the region of acceptance for the hypothesis testing (the model is not misspecified), and estimate the other parameters. The elaborate empirical application will be discussed in the next section.

6.4 Empirical application

In this section, we present the empirical results of a noise trader model with various FX data. First, we fix the system size (N) and the herding tendency (ν) parameter in the model while estimating the other parameters via SMM. Second, we present the case for estimating the trading volume parameters by fixing the herding tendency parameter. After finding a parameter region that can provide a reasonably good approximation to the data, we demonstrate how to decompose the return volatility into the news innovations and social interactions.

6.4.1 Data

For empirical application, we use various FX market data and compute the summary statistics: CHF/US, DM/US, FRF/US, GBP/US, JPY/US. Appendix B reports summary statistics of the data sets used in this study. All data range from

Jan.02.1973 to Dec.11.2009.⁹ The size of the sample is 9,000 at a daily frequency. We choose a 200 window length with 45 number of blocks. 1,000 samples are drawn at random using the bootstrap procedure.

6.4.2 Parameter calibration I: system size and herding tendency

In this section, we demonstrate that the appropriately chosen system size plays an important role in detecting the social interaction effects in this model. To see this, we fix the system size and attempt to estimate the other parameters; i.e. ν , $N_F T_F$, $N_C T_C$, δ_0 . In particular, we investigate the variation in the parameter estimates with respect to the system size. Note here that we report the results of this experiment using the DM/US data. In fact, the empirical results are robust to the other FX data.¹⁰ The following starting values are used for this estimation: $\nu = 0.5$, $N_F T_F = 10$, $N_C T_C = 1.0$, $100 \cdot \delta_0 = 0.5$.

Table 6.6: Variation in parameters estimates from system size: DM/US

N	parameter				critereon	5%	1%
	ν	$N_F T_F$	$N_C T_C$	$100 \cdot \delta_0$	Ψ	18.96	24.14
10	0.093	23.438	1.258	0.411	69.60		R
50	0.145	10.449	1.268	0.294	18.42		A
100	0.224	14.088	1.951	0.347	14.74		A
200	0.535	16.098	2.111	0.373	14.63		A
500	1.440	14.443	1.844	0.414	10.57		A
1000	2.352	15.854	2.287	0.440	10.30		A

Note: Statistical inference for model validation (R: rejected, A: accepted).

It is shown in Table 6.6 that the model using the large system size achieves a good approximation to the data; i.e. the objective function value is the smallest $\Psi=10.30$ for $N=1,000$. Indeed, it is noticeable that the appropriate number of noise traders play an important role in improving the fit of the model to the return volatility; the model with the small number of agents (i.e., $N = 10$) is rejected at 1% level as being a "true" model according to the block bootstrapping criterion. The empirical result shows that the total number of the noise traders can be set to the value, which is greater than 50. This experiment demonstrates that the trading volume and the herding tendency parameters are highly correlated; the more the interacting agents are employed for simulations, the larger values for the behavioral parameters (ν , $N_C T_C$) we obtain. In contrast, the estimated value for news innovation remains moderately stable between 0.30 and 0.44.

⁹The data were retrieved from the Sauder School of Business, University of British Columbia, <http://fx.sauder.ubc.ca/data.html>.

¹⁰To save space, we report the case of DM/US. The results of the system size remain almost the same across the other FX data.

Next, we set the system size and the herding tendency parameter to 100 and 0.5, respectively. From Table 6.6, we can see that the model with some of these parameter values can provide a reasonable approximation to the data; the variation in the parameter ν lies between 0.15 and 0.54 when the system size changes from 50 to 200. The model with these variations can be acceptable of a "correctly" specified case. Then we estimate the other parameters in the model: i.e. $N_F T_F$, $N_C T_C$, $100 \cdot \delta_0$. The parameter estimates are reported in Table 6.7. The following starting values are used for this estimation: $N_F T_F = 10$, $N_C T_C = 1.0$, $100 \cdot \delta_0 = 0.5$.

Table 6.7: Parameter estimates I of the noise trader model for FX data

	$N_F T_F$	$N_C T_C$	$100 \cdot \delta_0$	Ψ	inference
CHF/US	15.123	1.5190	0.422	14.16	A
DM /US	11.152	1.058	0.367	19.51	A
FRF/US	8.713	0.829	0.330	34.30	R
GBP/US	12.527	1.013	0.295	43.86	R
JPY/US	36.891	3.868	0.357	24.21	R

Note: Statistical inference for model validation (R: rejected, A: accepted).
The parameters N and ν are set to 100 and 0.5, respectively.

From the empirical results where we control the system size and the herding tendency parameter, it can be seen that the model achieves a good approximation to the data set of CHF and DM. On the other hand, the model is rejected when fitting the data set of FRF, GBP and JPY. Again, the estimate for the news innovation parameter appears stable lying between 0.30 and 0.43. More accurate estimates for news innovation and social interaction effects will be discussed later.

6.4.3 Parameter calibration II: trading volume

As it is discussed in the previous section, however, the obtained estimates are sensitive to the starting values due to the flatness and non-convexity in the objective function. To check this problem, we continue another experiment to find an acceptance region of the model. In this experiment, we set the total number of agents to 100 while the simulation size is set to 10. First, we fix the noise trading volume parameter to one and control the variation in the trading volume parameter for the fundamentalist; i.e., $N_F T_F = 1.0, 2.0, 5.0, 10.0, 15.0, 20.0, 30.0$. Second, we increase the noise trading volume parameter by two or five times (2 and 5) and repeat the estimation procedure. These experiment settings allow us to examine a model acceptance surface and its behavior according to the chosen parameter values.

Table 6.8 shows that we can find the parameter values ($N_F T_F$) between 5 and 15 for being non-rejected when the trading volume parameter for the noise trader is set to one ($N_C T_C = 1$).¹¹ The results show that the area of parameter space

¹¹To save space, we report the case of DM/US. The results of this experiment remain almost the same across the other FX data.

Table 6.8: Variation in parameter estimates from trading volume: DM/US

$N_C T_C$	label	$N_F T_F$						
		1.0	2.0	5.0	10.0	15.0	20.0	30.0
1.0	ν	0.022	0.049	0.161	0.474	0.580	0.642	0.575
	$100 \cdot \delta_0$	0.360	0.341	0.324	0.362	0.323	0.308	0.338
	obj.	2499.60	912.77	74.92	15.21	48.93	103.94	213.97
	inference	R	R	R	A	R	R	R
2.0	ν	n/a	0.014	0.041	0.106	0.278	0.484	0.587
	$100 \cdot \delta_0$	n/a	0.349	0.325	0.378	0.345	0.348	0.332
	obj.	n/a	880.36	97.45	16.19	14.97	23.86	46.91
	inference	n/a	R	R	A	A	A	R
5.0	ν	n/a	n/a	n/a	0.018	0.034	0.061	0.133
	$100 \cdot \delta_0$	n/a	n/a	n/a	0.369	0.347	0.348	0.345
	obj.	n/a	n/a	n/a	24.45	17.90	13.99	20.39
	inference	n/a	n/a	n/a	R	A	A	A

Note: Statistical inference for model validation (R: rejected, A: accepted). The notation n/a means that estimates are not available, because they hit a boundary point (negative value for the parameter ν).

for non-rejection shifts to the right when $N_C T_C$ increases; i.e., $5 < N_F T_F < 30$ for $N_C T_C = 2$, while $10 < N_F T_F$ for $N_C T_C = 5$. Moreover, if the numerical value for the trading volume parameter has changed a bit in simulations, we obtain different estimates for the herding tendency parameter (ν). Again, the estimated value for news innovation remains stable between 0.32 and 0.38. From this experiment, we found that the market volatility can be robustly decomposed into news innovation and social interactions, but the parameter, which controls the switching process, is highly correlated with the trading volume parameters.

6.4.4 Decomposition of return volatility: news innovation and social interactions

The results of the empirical applications in the previous section show that the trading volume parameters are correlated to the herding tendency parameter in the model; see Table 6.6 and 6.8. To cope with the identification problems in the group behavior of this model, we fix the trading volume parameters and attempt to estimate the magnitude of news innovation and social interactions. More generally, we set the parameters $N_F T_F$ and $N_C T_C$ to 10 and unity, respectively. This numerical value can be a reasonable choice for our current study, because this value provides upper and lower bounds on the price path. Since the opinion index is multiplied by the relative ratio of trading volume parameters for the noise traders and fundamentalists ($\frac{N_C T_C}{N_F T_F} = \frac{1}{10}$), the price dynamics in Equation (6.8) are driven by changes in fundamentals and the the opinion index; i.e. 1 ± 0.1 given that we set the logarithmic value for the baseline fundamental value to one in our artificial economy.

Table 6.9: Parameter estimates II of the noise trader model for FX data

	ν	$100 \cdot \delta_0$	Ψ	inference
CHF/US	0.543	0.421	15.09	A
DM /US	0.463	0.364	12.49	A
FRF/US	0.537	0.313	35.67	R
GBP/US	0.324	0.304	25.33	A
JPY/US	0.509	0.353	31.04	R

Note: Statistical inference for model validation (R: rejected, A: accepted).

The parameters N_{FTF} and N_{CTC} are set to 10 and unity, respectively.

It is shown in Table 6.9 that the model fits the CHF, DM, and GBP data sufficiently well, whereas its fit to the FRF and JPY data is clearly not so good; i.e. see the estimated loss function values. We can see that the DM and CHF markets ($100 \cdot \hat{\delta}_0 = 0.421, 0.364$) are more affected by the news arrivals than the others. Concerning the estimated value for the velocity of the switching process, we can infer that the noise traders ($\hat{\nu} = 0.543, 0.537, 0.509$) are more active in the CHF, FRF and JPY markets than the others.

Table 6.10: Decomposition of stochastic processes for FX data

label	std.	news innovation ($100 \cdot \hat{\delta}_0$)	social interactions ($\hat{\nu}$)
CHF/US	0.734	0.421 (57.36%)	42.64%
DM /US	0.651	0.364 (55.91%)	44.09%
FRF/US	0.636	0.313 (49.21%)	50.79%
GBP/US	0.592	0.304 (51.35%)	48.65%
JPY/US	0.657	0.353 (53.73%)	46.27%

Note: "std." denotes the standard deviation of returns.

Further, Table 6.10 illustrates that the market volatility of various FX markets can be robustly decomposed into two parts: impacts of news innovation (45-55%) and social interactions (45-55%). Note that under these parameters, the model predicts that the active participation of the noise traders in the market give rise to volatility clustering and market turbulence. Put differently, erratic price movements are commonly associated with social interactions of the noise traders in the market. This prediction is in fact verified empirically. These observations are illustrated in Figure 6.10. The parameters used in the simulation are $\nu = 0.463$, $N_{FTF} = 10.0$, $N_{CTC} = 1.0$ and $100 \cdot \delta_0 = 0.364$. We also simulate an absolute value for the excess demands of the noise traders ($ED_{Nc} = |N_{CTC} \cdot x_t|$); when the noise traders buy or sell a large amount of assets in the marketplace, we often observe the volatile price movements.

Overall, we found minimum values for the objective function (15.09, 12.49 and

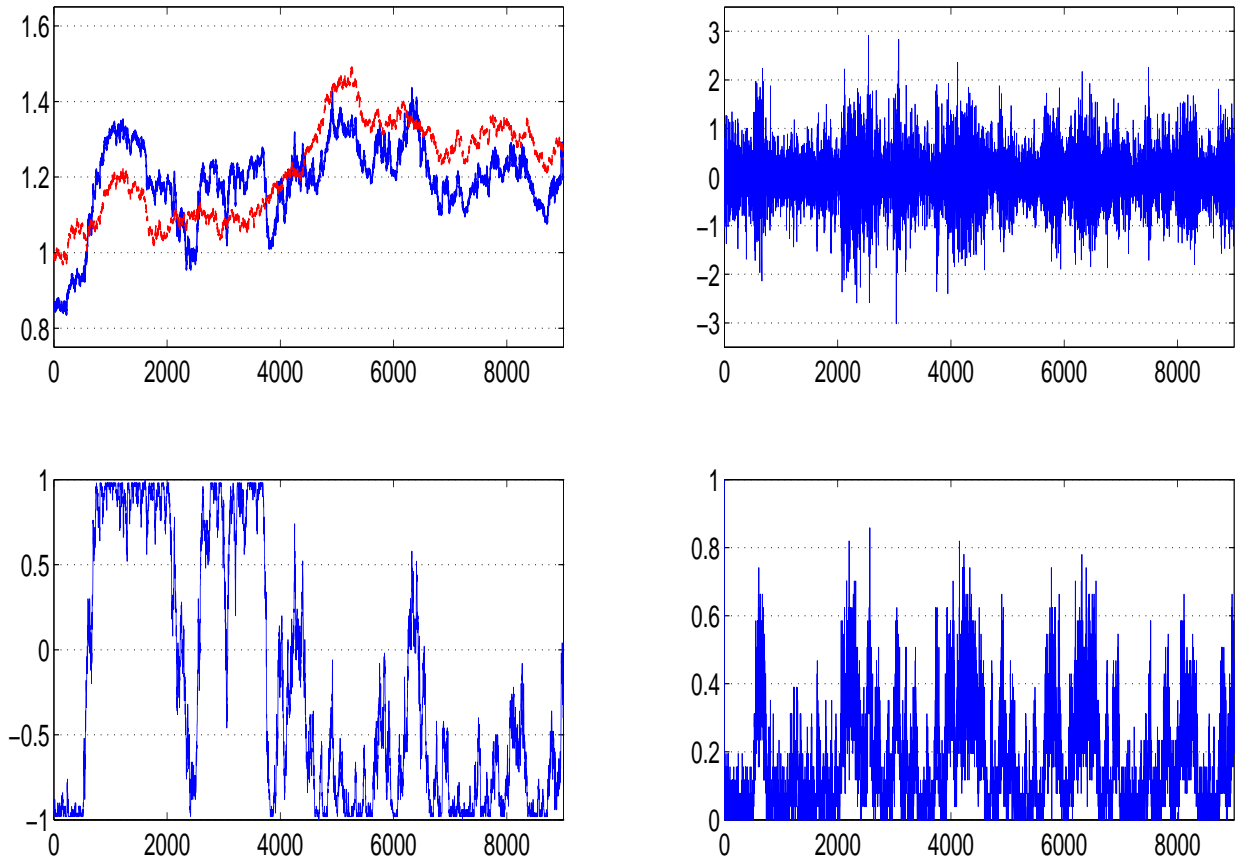


Figure 6.10: Model-generated data using the estimated parameters

Note: The upper panels plot prices (left) and returns (right) while the lower panels plot the corresponding opinion index (left) and excess demands of the noise traders (right).

25.33) using the CHF, DM, GBP data. Since these values do not exceed the criterion of model misspecification from the block bootstrap method, we can not reject the null hypothesis that the characteristics of the empirical data are matched by the model-generated ones with the parameter estimates in Table 6.9. Indeed, Table 6.11 shows that the model-generated moments match the empirical counterparts almost perfectly. However, the objective function values for the FRF and JPY market data do exceed the criteria from the block bootstrap method at 1% level ($35.67 > 24.10$, $31.04 > 25.89$); the model is misspecified.

Table 6.11: Moment estimates of FX data

	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
CHF	-0.0133	0.5382	0.0195	0.1365	0.1247	0.1111	0.0916	0.0712	0.9142	3.4123
NT	0.0003	0.4562	0.0266	0.1434	0.1347	0.1230	0.1058	0.0730	0.9122	3.9418
DM	-0.0089	0.4241	0.0287	0.1386	0.1453	0.1253	0.0951	0.0799	0.8900	3.3781
NT	0.0003	0.4090	0.0337	0.1502	0.1431	0.1361	0.1213	0.0639	0.9261	3.8753
FRF	-0.0007	0.4040	0.0265	0.1316	0.1539	0.1317	0.1018	0.0946	0.8881	3.1911
NT	0.0003	0.3724	0.0440	0.1572	0.1486	0.1413	0.1262	0.0666	0.9236	3.8550
GBP	0.0032	0.3507	0.0525	0.1716	0.1759	0.1566	0.1338	0.0671	0.9197	3.1955
NT	0.0001	0.3343	0.0374	0.1806	0.1687	0.1557	0.1380	0.0870	0.9013	3.7778
JPY	-0.0120	0.4311	0.0222	0.1788	0.1471	0.1165	0.1143	0.1297	0.8693	3.3463
NT	0.0002	0.3880	0.0282	0.1667	0.1516	0.1364	0.1154	0.0900	0.8935	3.8796

6.5 Conclusion

In this chapter, we proposed the moment-based estimation to examine the effects of the group behavior on an opinion formation process in financial markets. In a simple interacting agent model, the transition rates of the opinion dynamics affect the price movements; we assume that fundamentals are described by a Brownian motion. However, the mapping between the model parameters and the objective function is non-convex and flat with respect to the behavioral parameters during the estimation. To circumvent identification problems in the model, we first calibrated the system size and trading volume parameters, and used simulation based inference to validate the model as being a "true" one. Accordingly, the SMM estimation provides the model parameter estimates by matching (as closely as possible) the moments of the FX data; the market volatility can be robustly decomposed into the news innovation and social interactions.

As Keynes (1936) wrote about an imaginary beauty contest in his famous work *General Theory of Employment Interest and Money*, investors make a rule of thumb estimate of the market sentiments. According to their assessment of the market situation, they are likely to buy or sell assets out of speculative motive. Because of this, we can establish a connection between the effects of social interactions and the price dynamics in the market. Indeed, his metaphor for the group behavior is empirically verified by taking a simple interaction-based model to the FX data. This implies that the behavioral analysis of heterogeneous investors can successfully reproduce the empirical feature of the financial data. The empirical results also show that the transition rates of the historical market sentiment give rise to excessive volatility in financial markets.

According to a test criterion from the block bootstrap resample, we cannot reject the null hypothesis that the empirical characteristic of some FX data can be explained by a simple interacting agent model. However, we found that the model

is misspecified for the FRF and JPY data. Since the moment-based estimation provides benchmark values for the behavioral parameter in a *Markov chain* of the opinion dynamics, we can further continue our study of the model identifications using the simulation based approaches such as a *Markov chain Monte Carlo* and filtering methods. Future research is needed to examine the variation in the behavioral parameter estimates in our study.

Appendix A: Choice of moments

We assume that the data sets used here are second-order stationary. Then we denote by μ_r , $\mu_{|r|}$, and μ_{r^2} first moments of the population for the raw, absolute and square returns respectively. The second moments of r , $|r|$, and r^2 are respectively σ_r , $\sigma_{|r|}$, and σ_{r^2} . Autocorrelation coefficient ρ_τ refers the correlation of the return series with its own past values after time lag τ . The moment conditions are chosen among the following 11 moments ($m_1 - m_{11}$):

$$m_1 = E[r_t] - \mu_r$$

$$m_2 = E[r_t^2] - \sigma_r^2$$

$$m_3 = E\left[\frac{(r_t - \mu_r)(r_{t-1} - \mu_r)}{\sigma_r^2}\right] - \rho_{1,r}$$

$$m_{3+i} = E\left[\frac{(|r_t| - \mu_{|r|})(|r_{t-\tau}| - \mu_{|r|})}{\sigma_{|r|}^2}\right] - \rho_{\tau,|r|}, \quad i=1, 2, 3, 4 \quad \& \quad \tau = 1, 5, 10, 20$$

The parameters ($m_{ID,1} - m_{ID,3}$) of the auxiliary models (the univariate GARCH (1,1) model and the Hill estimator) are estimated via the indirect estimation method:

GARCH (1,1)

$$r_t = \gamma_0 + \varepsilon_t, \quad \varepsilon_t \sim N(0, h_t)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

Tail index; see Hill (1975).

$$\xi = \frac{1}{k-1} \sum_{i=1}^{k-1} \ln r_{i,n} - \ln r_{k,n}, \quad \text{for } k \geq 2$$

$$m_{ID,1} = \alpha_1$$

$$m_{ID,2} = \beta_1$$

$$m_{ID,3} = \alpha \quad (:= \frac{1}{\xi})$$

B: Summary statistics of FX data

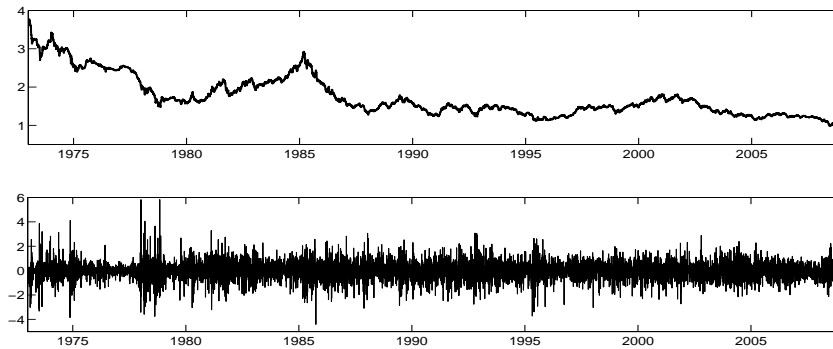


Figure 6.11: Prices (upper) and returns (lower) of CHF/US data

Table 6.12: Statistics of empirical moments and block bootstrap samples: CHF/US

Label	Statistic	Empirical	Bootstrap
m_1	$E[r_t]$	-0.0133	-0.0136
m_2	$\text{Var}[r_t]$	0.5382	0.5392
m_3	$\text{corr}(r_t, r_{t+1})$	0.0195	0.0182
m_4	$\text{corr}(r_t , r_{t+1})$	0.1365	0.1340
m_5	$\text{corr}(r_t , r_{t+5})$	0.1247	0.1203
m_6	$\text{corr}(r_t , r_{t+10})$	0.1111	0.1099
m_7	$\text{corr}(r_t , r_{t+20})$	0.0916	0.0835
$m_{ID,1}$	α_1 (GARCH(1,1))	0.0712	0.0754
$m_{ID,2}$	β_1 (GARCH(1,1))	0.9142	0.9042
$m_{ID,3}$	α (Right tail, 5 - 10 %)	3.4123	3.4429
	Test criterion	5%: 18.42	1%: 23.44

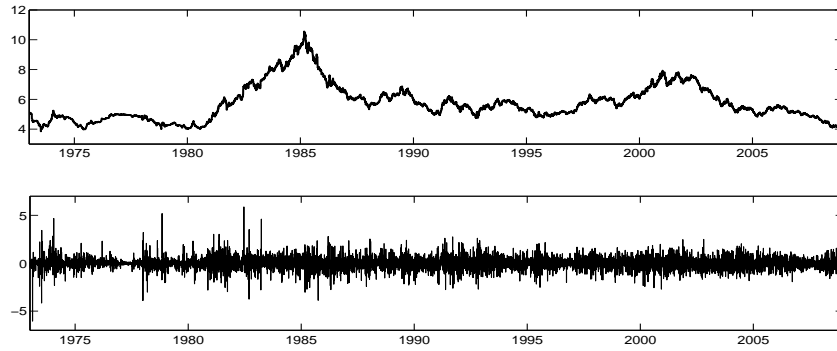


Figure 6.12: Prices (upper) and returns (lower) of FRF/US data

Table 6.13: Statistics of empirical moments and block bootstrap samples: FRF/US

Label	Statistic	Empirical	Bootstrap
m_1	$E[r_t]$	-0.0007	-0.0006
m_2	$\text{Var}[r_t]$	0.4040	0.4037
m_3	$\text{corr}(r_t, r_{t+1})$	0.0265	0.0252
m_4	$\text{corr}(r_t , r_{t+1})$	0.1316	0.1299
m_5	$\text{corr}(r_t , r_{t+5})$	0.1539	0.1473
m_6	$\text{corr}(r_t , r_{t+10})$	0.1317	0.1271
m_7	$\text{corr}(r_t , r_{t+20})$	0.1018	0.0973
$m_{ID,1}$	α_1 (GARCH(1,1))	0.0946	0.0935
$m_{ID,2}$	β_1 (GARCH(1,1))	0.8881	0.8866
$m_{ID,3}$	α (Right tail, 5 - 10 %)	3.1911	3.2142
	Test criterion	5%: 18.24	1%: 24.10

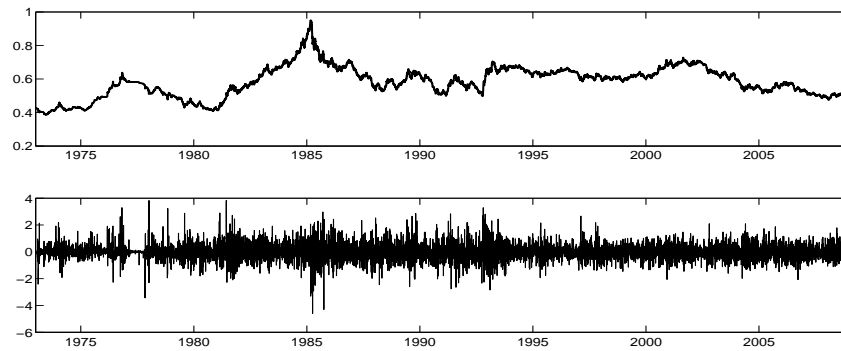


Figure 6.13: Prices (upper) and returns (lower) of GBP/US data

Table 6.14: Statistics of empirical moments and block bootstrap samples:GBP/US

Label	Statistic	Empirical	Bootstrap
m_1	$E[r_t]$	0.0032	0.0032
m_2	$\text{Var}[r_t]$	0.3507	0.3519
m_3	$\text{corr}(r_t, r_{t+1})$	0.0525	0.0516
m_4	$\text{corr}(r_t , r_{t+1})$	0.1716	0.1690
m_5	$\text{corr}(r_t , r_{t+5})$	0.1759	0.1718
m_6	$\text{corr}(r_t , r_{t+10})$	0.1566	0.1547
m_7	$\text{corr}(r_t , r_{t+20})$	0.1338	0.1281
$m_{ID,1}$	α_1 (GARCH(1,1))	0.0671	0.0796
$m_{ID,2}$	β_1 (GARCH(1,1))	0.9197	0.8939
$m_{ID,3}$	α (Right tail, 5 - 10 %)	3.1955	3.2307
	Test criterion	5%: 18.77	1%: 25.89

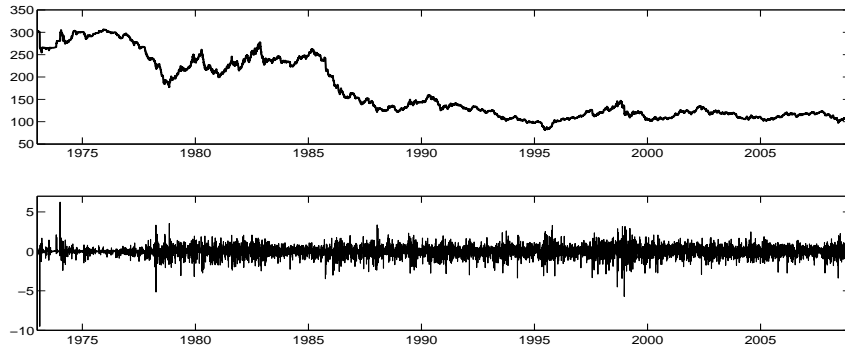


Figure 6.14: Prices (upper) and returns (lower) of JPY/US data

Table 6.15: Statistics of empirical moments and block bootstrap samples: JPY/US

Label	Statistic	Empirical	Bootstrap
m_1	$E[r_t]$	-0.0120	-0.0122
m_2	$\text{Var}[r_t]$	0.4311	0.4301
m_3	$\text{corr}(r_t, r_{t+1})$	0.0222	0.0218
m_4	$\text{corr}(r_t , r_{t+1})$	0.1788	0.1740
m_5	$\text{corr}(r_t , r_{t+5})$	0.1471	0.1420
m_6	$\text{corr}(r_t , r_{t+10})$	0.1165	0.1045
m_7	$\text{corr}(r_t , r_{t+20})$	0.1143	0.0993
$m_{ID,1}$	α_1 (GARCH(1,1))	0.1297	0.1820
$m_{ID,2}$	β_1 (GARCH(1,1))	0.8693	0.7414
$m_{ID,3}$	α (Right tail, 5 - 10 %)	3.3463	3.3674
	Test criterion	5%: 18.22	1%: 26.58

General Conclusion

In this thesis, we empirically examined the importance of the behavior of economic agents for analyzing the macroeconomic and financial market models. In particular, we evaluated empirical performance of the DSGE models and ABMs by using a formal model comparison method. The empirical results show that the behavioral parameters of these models can be estimated by the method of moments estimation and simulation based inference. Accordingly we can draw conclusions that the group behavior in the economy can be one of the main driving forces behind the macroscopic dynamics in macroeconomic and financial data. In this way, we challenged the theory of the representative agent by showing that behavioral heterogeneity and social interactions can provide a good approximation to real data.

Main findings of this thesis and future research can be summarized as follows:

(1) In chapter 2, we estimate the parameters of a small-scale DSGE model by using the method of moments estimation and contrast them with the Bayesian estimation. In particular, we attempt to compare two specifications of the model using a formal test: forward- versus backward-looking behavior in the New-Keynesian Phillips curve. Our findings suggest that the moment-based estimation provides evidence for backward-looking behavior in the historical macro data of the US. In addition, the results of the model comparison show that a hybrid version of the New-Keynesian model has a better fit to the data than a model with forward-looking expectations for the *Great Inflation* period. A potentially interesting future research topic would be to compare the empirical performance of the DSGE model with a parsimonious vector autoregressive model, since the latter is extensively used for economic analysis.

(2) In chapter 3, the method of moments and maximum likelihood estimation are used to examine the significance of a lagged term in the New-Keynesian Phillips curve and the IS equation. Because of this, we can establish the empirical relationship between the price-setting/consumption behavior and the sources of persistence in inflation and output; the inclusion of backward-looking behavior improves the fit of the model while offsetting the influence of inherited and extrinsic persistence. Moreover, it is shown that intrinsic persistence plays a important role for capturing the structural features of inflation and output. We also carry out Monte Carlo experiments to examine the validity of the chosen moment conditions and the finite sample properties of the estimation methods. The empirical performance of model

selection methods is discussed along the lines of the Akaike's and the Bayesian information criterion.

(3) In chapter 4, we empirically examine the model of De Grauwe (2011) in which agents have cognitive limitations. This model incorporates the discrete choice theory into a monetary DSGE framework where agent's belief can display endogenous waves of market optimism and pessimism. Especially, we estimate the behavioral parameters governing market expectations of heterogeneous agents using the method of moment estimator. Since the model is found to be nonlinear in the switching mechanism, we cannot easily obtain analytical moment conditions of the system for estimation. Therefore, we use a simulation based approach to approximate the moment conditions. The fit of the model via the "Moment Matching" shows that the agents had expected moderate deviations of the output gap from its steady state value with low uncertainty. This provides an empirical evidence for backward-looking behavior in the expectation formation process. Since the estimation of a small-scale DSGE model is considered to be satisfactory, we can further continue the model estimation with much richer models like e.g. a medium-scale version of the DSGE model developed by Smets and Wouters (2003, 2005). We leave this issue to future research.

(4) In chapter 5, first, we study the model dynamics of the adaptive belief system of Gaunersdorfer and Hommes (2007) and estimate the parameters of the model with two types of noise term: additive and multiplicative. The empirical results show that this model cannot explain the empirical regularities of financial data (i.e. the absence of long range dependence in the return volatility). Indeed, the model's fit of the S&P 500 data is dependent on the noise specification in the dynamics. Second, we present a structural stochastic volatility model of Franke and Westerhoff (2012) and estimate the parameters of two model specifications: a herding and a wealth mechanism. Then we use the formal model comparison method to assess the significance of the fit between the two models. As a result, we find that the two trading mechanisms do not play a significant role in approximating the return volatility. The inconclusive result can be further examined by verifying the power of the test and developing statistical approach under a framework of the misspecified models.

(5) In chapter 6, we study the group behavior of a noise trader model of Alfarano and Lux (2007). By applying the simulated method of moments estimator to the model, we can examine the effects of the group behavior on financial data. However, the stochastic switching rule gives rise to multimodalities in the parameter space; this prevents a standard econometric analysis from inferring the exact time-dependent moments of data. To circumvent these problems, we calibrate values for the system size and trading volume parameters by allowing for the social interaction effects in simulations. We use a large number of agents (around 100) and trading volume parameters in order to reproduce the empirical regularities observed in various FX data. Afterwards, we estimate the effects of the news innovations and social in-

teractions on the data. The empirical findings of this chapter suggest that return volatility in financial markets under scrutiny can be robustly decomposed into the news innovations (45-55%) and social interactions (45-55%). Since the moment-based estimation provides benchmark values for the parameter of a *Markov chain* of the opinion dynamics, we can further continue our study of the model identifications using the simulation based approaches such as a *Markov chain Monte Carlo* and filtering methods. Future research is needed to examine the variation in the behavioral parameter estimates in our study.

In this thesis, our intention in examining the non-linear and stochastic economic models was to provide a practical guidance to those who develop, evaluate, and apply complex economic models. Since complexity in these models hinders a direct empirical application, we showed that the moment-based estimation can provide reasonable values for the behavioral parameters, and the fit of the complex models can be evaluated along the lines of chosen moment conditions. Therefore the statistical methods discussed herein can be applicable to the estimation of several variants of the structural macroeconomic models and the agent-based models. From this, we can find an empirical connection between the behavior of economic agents and the model dynamics. Future research is needed to investigate the properties of the moment conditions for these complex models and statistical inference.

References

Reference to Chapter 1

ALFARANO, S. AND LUX, T. (2007): A Noise Trader Model as a Generator of Apparent Financial Power Laws and Long Memory, *Macroeconomic Dynamics*, Vol. 11, pp. 80–101.

DE GRAUWE, P. (2011): Animal Spirits and Monetary Policy. *Economic Theory*, Vol. 47, pp. 423–457.

DUFFIE, D. AND SINGLETON, K. (1993): Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, Vol. 61(4), pp. 929–52.

FRANKE, R., AND WESTERHOFF, F. (2012): Converse Trading Strategies, Intrinsic Noise and the Stylized Facts of Financial Markets. *Quantitative Finance*, Vol. 12(3), pp. 425–436.

GAUNERSDORFER, A. AND HOMMES, C. (2007): A Nonlinear Structural Model for Volatility Clustering. In: Teyssiere, G. and Kirman, A. (ed.), *Long Memory in Economics*, Springer, pp. 265-288.

HANSEN, L.P. (1982): Large sample properties of generalized method of moments estimators. *Econometrica*, Vol. 50, pp. 1029–1054.

HNATKOVSKA, V. AND MARMER, V. AND TANG, Y. (2012): Comparison of Misspecified Calibrated Models: the Minimum Distance Approach. *Journal of Econometrics*, *Forthcoming*.

McFADDEN, D. (1989): A method of simulated moments for estimation of discrete choice models without numerical integration. *Econometrica*, Vol. 57, pp. 995–1026.

NEWKEY W. AND McFADDEN, D. (1994): *Large sample estimation and hypothesis testing*. in Handbook of Econometrics, Elsevier Science, Ch.36, pp. 2111–2245.

PAKES, A. AND POLLARD, D. (1989): Simulation and the asymptotics of optimization estimators. *Econometrica*, Vol. 57(5), pp. 1027–1057.

WINKER, P. AND GILLI M. AND JELESKOVIC V. (2007): An Objective Function for Simulation Based Inference on Exchange Rate Data. *Journal of Economic Interaction and Coordination*, Vol. 2, pp. 125–145.

Reference to Chapter 2

ADOLFSON, M., LASEEN, S., LINDÉ, J. AND VILLANI, M. (2007): Bayesian estimation of an open economy DSGE model with incomplete pass-through. *Journal of International Economics*, 72, 481–511.

ALTIG, D., CHRISTIANO, L.J., EICHENBAUM, M. AND LINDÉ, J. (2011): Firm-specific capital, nominal rigidities and the business cycle. *Review of Economic Dynamics*, 14, 225–247.

ALTONJI, J. AND SEGAL, L. (1996): Small-sample bias in GMM estimation of covariance structures. *Journal of Business and Economic Statistics*, 14, 353–366.

AMBLER, S., GUAY, A. AND PHANEUF, L. (2011): Endogenous business cycle propagation and the persistence problem: The role of labor-market frictions. *Journal of Economic Dynamics and Control* (in press).

AVOUYI-DOVI, S. AND MATHERON, J. (2007): Technology shocks and monetary policy: Revisiting the Fed's performance. *Journal of Money, Credit and Banking*, 39, 471–507.

BAI, Z.-Z., GUO, X.-X. AND YIN, J.-F. (2005): On two iteration methods for the quadratic matrix equations. *International Journal of Numerical Analysis and Modeling*, 2, Supp, 114–122.

BENATI, L. AND SURICO, P. (2009): VAR analysis and the Great Moderation. *American Economic Review*, 99, 1636–1652.

BERNANKE, B. AND MIHOV, I. (1998): Measuring monetary policy. *Quarterly Journal of Economics*. 113, 869–902.

BEYER, A. AND FARMER, R.E.A. (2007): Testing for indeterminacy: An application to U.S. monetary policy: Comment. *American Economic Review*, 97, 524–29.

BINDER, M. AND PESARAN, M.H. (1996): Multivariate rational expectations models and macroeconomic modeling: A review and some new results. In: Pesaran, M.H. and Wickens, M. (Eds.), *Handbook of Applied Econometrics: Macroeconomics*. Oxford: Blackwell; pp. 139–187.

BOIVIN, J. AND GIANNONI, M.P. (2006): Has monetary policy become more effective? *Review of Economics and Statistics*, 88, 445–462.

BROOKS, S.P. AND GELMAN, A. (1998): General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*,

7, 434–455.

CARRASCO, M. AND FLORENS, J.-P. (2002): Simulation-based method of moments and efficiency. *Journal of Business and Economic Statistics*, 20, 482–492.

CASTELNUOVO, E. (2010): Trend inflation and macroeconomic volatilities in the post WWII U.S. economy. *North American Journal of Economics and Finance*, 21, 19–33.

CHO, S. AND MORENO, A. (2006): A small-sample study of the New Keynesian macro model. *Journal of Money, Credit and Banking*, 39, 471–507.

CHRISTIANO, L.J., EICHENBAUM, M. AND EVANS, C.L. (2005): Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113, 1–45.

COCHRANE, J.H. (2001): *Asset Pricing*. Princeton: Princeton University Press.

COGLEY, T., PRIMICERI, G.E. AND SARGENT, T.J. (2010): Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics*, 2:1, 43–69.

COGLEY, T. AND SBORDONE, A. (2008): Trend inflation, indexation, and inflation persistence in the New Keynesian Phillips curve. *American Economic Review*, 98, 2101–26.

COLLARD, F., FÈVE, P., LANGOT, F. AND PERRAUDIN, C. (2002): A structural model of U.S. aggregate job flows. *Journal of Applied Econometrics*, 17, 197–223.

DAVIDSON, R. AND MACKINNON, J.G. (2004), *Econometric Theory and Methods*. Oxford: Oxford University Press.

DEL NEGRO, M. AND SCHORFHEIDE, F. (2004): Priors from general equilibrium models for VARS. *International Economic Review*, 45, 643–673.

DEL NEGRO, M., SCHORFHEIDE, F., SMETS, F. AND WOUTERS, R. (2007): On the fit of New Keynesian models. *Journal of Business and Economic Statistics*, 25, 123–143.

FÈVE, P., MATHERON, J. AND SAHUC, J.-G. (2009): Inflation target shocks and monetary policy inertia in the euro area. *The Economic Journal*, 120, 1100–1124.

FUHRER, J. (2006): Intrinsic and inherited inflation persistence. *International Journal of Central Banking*, 2:3, 49–86.

GILL, P., MURRAY, W. AND WRIGHT, M. (1981): *Practical Optimization*. New York: Academic Press.

GORODNICHENKO, Y. AND NG, S. (2010): Estimation of DSGE models when the data are persistent. *Journal of Monetary Economics*, 57, 325–340.

HAIRAULT, J., LANGOT, F. AND PORTIER, F. (1997): Time to implement and aggregate fluctuations. *Journal of Economic Dynamics and Control*, 22, 109–121.

HENZEL, S., HÜLSEWIG, O., MAYER, E. AND WOLLMERSHÄUSER, T. (2009): The price puzzle revisited: Can the cost channel explain a rise in inflation after a monetary policy shock? *Journal of Macroeconomics*, 31, 268–289.

HNATKOVSKA, V., MARMER, V. AND TANG, Y. (2009): Comparison of misspecified calibrated models: The minimum distance approach. Working Paper, University of British Columbia.

HÜLSEWIG, O., MAYER, E. AND WOLLMERSHÄUSER, T. (2009): Bank behavior, incomplete interest rate pass-through, and the cost channel of monetary policy transmission. *Economic Modelling*, 26, 1310–1327.

IRELAND, P.N. (2007): Changes in the Federal Reserve's inflation target: Causes and consequences. *Journal of Money, Credit and Banking*, 39, 1851–82.

JONSSON, G. AND KLEIN, P. (1996): Stochastic fiscal policy and the Swedish Business Cycle. *Journal of Monetary Economics*, 38, 245–268.

KARAMÉ, F., PATUREAU, L. AND SOPRASEUTH, T. (2008): Limited participation and exchange rate dynamics: Does theory meet the data? *Journal of Economic Dynamics and Control*, 32, 1041–1087.

KIM, J. AND RUGE-MURCIA, F.J. (2011): Monetary policy when wages are downwardly rigid: Friedman meets Tobin. *Journal of Economic Dynamics and Control* (in press).

LAUBACH, T. AND WILLIAMS, J.C. (2003): Measuring the natural rate of interest. *The Review of Economics and Statistics*, 85, 1063–1070.

LINDÉ, J. (2005): Estimating New-Keynesian Phillips curves: A full information maximum likelihood approach. *Journal of Monetary Economics*, 52, 1135–49.

LUBIK, T. AND SCHORFHEIDE, F. (2004): Testing for indeterminacy: An application to U.S. monetary policy. *American Economic Review*, 94, 190–217.

LUBIK, T. AND SCHORFHEIDE, F. (2007): Do central banks respond to exchange rate movements? A structural investigation. *Journal of Monetary Economics*, 54, 1069–1087.

LÜTKEPOHL, H. (2007): *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.

MATHERON, J. AND POILLY, C. (2009): How well does a small structural model with sticky prices and wages fit postwar U.S. data? *Economic Modelling*, 26, 266–284.

MESONNIER, J.-S. AND RENNE, J.-P. (2007): A time-varying ‘natural’ rate of interest for the euro area. *European Economic Review*, 51, 1768–1784.

NEWKEY, W.K. AND MCFADDEN, D. (1994): Large sample estimation and hypothesis testing. In: Engle, R. and McFadden, D. (Eds.), *Handbook of Econometrics*, Vol. 4, Amsterdam: North Holland; pp. 2113–2245.

ROTEMBERG, J.J. AND WOODFORD, M. (1997): An optimization-based econometric framework for the evaluation of monetary policy. In: Bernanke, B. and Rotemberg, J.J. (Eds.), *NBER Macroeconomics Annual*. Cambridge, MA: MIT Press.

RUDD, J. AND WHELAN, K. (2005): Modelling inflation dynamics: A critical review of recent research. Federal Reserve Board, Washington, D.C., Finance and Economics Discussion Series, 2005–66.

RUDD, J. AND WHELAN, K. (2007): Modeling inflation dynamics: A critical review of recent research. *Journal of Money, Credit and Banking*, Supplement to Vol. 39, No. 1, 155–170.

RUDEBUSCH, G.D. (2002): Term structure evidence on interest rate smoothing and monetary policy inertia. *Journal of Monetary Economics*, 49, 1161–87.

SALEMI, M.K. (2006): Econometric policy evaluation and inverse control. *Journal of Money, Credit, and Banking*, 38, 1737–64.

SCHORFHEIDE, F. (2005): Learning and monetary policy shifts. *Review of Economic Dynamics*, 8, 392–419.

SCHORFHEIDE, F. (2008): DSGE model-based estimation of the New Keynesian Phillips curve. *Federal Reserve Bank of Richmond Economic Quarterly*, 94, 397–433.

SHEEDY, K.D. (2010): Intrinsic inflation persistence. *Journal of Monetary Economics*, 57, 1049–61.

SMETS, F. AND WOUTERS, R. (2003): An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, 1, 1123–1175.

SMETS, F. AND WOUTERS, R. (2007): Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, 97, 586–606.

TAYLOR, J.B. (1993): Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.

Reference to Chapter 3

- ALTONJI, J. AND SEGAL, L. (1996): Small-sample bias in GMM estimation of covariance structures. *Journal of Business and Economic Statistics*, 14, 353–366.
- AMATO, J.D. AND LAUBACH, T. (2003): Rule-of-thumb behaviour and monetary policy. *European Economic Review*, Vol. 47(5), pp. 791–831.
- AMATO, J.D. AND LAUBACH, T. (2004): Implications of habit formation for optimal monetary policy. *Journal of Monetary Economics*, Vol. 51, pp. 305–325.
- ANATOLYEV, S. AND GOSPODINOV, N. (2011): *Methods for Estimation and Inference in Modern Econometrics*. New York: A Chapman and Hall Book.
- BINDER, M. AND PESARAN, H. (1995): Multivariate Rational Expectations Models and Macroeconomic Modeling: a Review and Some Insights, In: Pesaran, M.H. and Wickens, M. (ed.), *Handbook of Applied Econometrics*. Oxford: Basil Blackwell, pp. 139–187.
- CALDARA, D., FERNANDEZ-VILLAVARDE, J., RUBIO-RAMIREZ, J. AND YAO, W. (2012): Computing DSGE models with recursive preferences and stochastic volatility. *Review of Economic Dynamics*, Vol. 15(2), pp. 188–206.
- CANOVA, F. AND SALA, L. (2009): Back to square one: identification issues in DSGE models. *Journal of Monetary Economics*, Vol. 56(4), pp. 431–449.
- CARRILLO, J.A., FÉVE, P. AND MATHERON J. (2007): Monetary policy inertia or persistent shocks: a DSGE analysis. *International Journal of Central Banking*, Vol. 3(2), pp. 1–38.
- CASTELNUOVO, E. (2010): Trend inflation and macroeconomic volatilities in the Post WWII U.S. economy. *North American Journal of Economics and Finance*, Vol. 21, pp. 19–33.
- CHRISTIANO, L., EICHENBAUM, M. AND C. EVANS (2005): Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* Vol. 113(1), pp. 1–45.
- COENEN, G. (2005): Asymptotic confidence bands for the estimated autocovariance and autocorrelation functions of vector autoregressive models. *Empirical Economics*, Vol. 30(1), pp. 65–75.
- COGLEY, T. AND SBORDONE, A. (2008): Trend inflation, indexation, and inflation persistence in the New-Keynesian Phillips curve. *American Economic Review*, Vol.

98(5), pp. 2101–2126.

COGLEY, T., PRIMICERI, G.E. AND SARGENT, T.J. (2010): Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics*, Vol. 2(1), pp. 43–69.

DE GRAUWE, P. (2011): Animal spirits and monetary policy. *Economic Theory*, Vol. 47, pp. 423–457.

DRIDI, R., GUAY, A. AND RENAULT, E. (2007): Indirect inference and calibration of dynamic stochastic general equilibrium models. *Journal of Econometrics*, Vol. 136(2), pp. 397–430.

FRANKE, R., JANG, T.-S. AND SACHT, S. (2011): *Moment matching versus Bayesian estimation: backward-looking behaviour in the New-Keynesian three-equations model*. Economic Working paper 2011-10, University of Kiel.

FUHRER, J.C. (1997): The (un)importance of forward-Looking behavior in price specifications. *Journal of Money, Credit and Banking*, Vol. 29(3), pp. 338–350.

FUHRER, J.C. (2000): Habit formation in consumption and its implications for monetary policy models. *American Economic Review*, Vol. 90(3), pp. 367–390.

GALI, J. AND GERTLER, M. (1999): Inflation dynamics: a structural econometric analysis. *Journal of Monetary Economics*, Vol. 44, pp. 195–222.

GILL, P., MURRAY, W. AND WRIGHT, M. (1981): *Practical Optimization*. Academic Press.

GOLDEN, R.M. (2000): Statistical tests for comparing possibly misspecified and nonnested models. *Journal of Mathematical Psychology*, Vol. 44, pp. 153–170.

GOLDEN, R.M. (2003): Discrepancy risk model selection test theory for comparing possibly misspecified or nonnested models. *Psychometrika*, Vol. 68(2), pp. 229–249.

GOURIEROUX, C. AND MONFORT, A. (1995): Testing, encompassing and simulating dynamic econometric models. *Econometric Theory*, Vol. 11, pp. 195–228.

GREGORY, A.W. AND SMITH, G.W. (1991): Calibration as testing: inference in simulated macroeconomic models. *Journal of Business and Economic Statistics*, Vol. 9(3), pp. 297–303.

HALL, A.R., INOUE, A., NASON, J.M. AND ROSSI, B. (2012): Information criteria for impulse response function matching estimation of DSGE models, *Journal of*

Econometrics, Vol. 170(2), pp. 499–518.

HNATKOVSKA, V., MARMER, V. AND TANG, Y. (2012): Comparison of misspecified calibrated models: the minimum distance approach. *Journal of Econometrics*, Vol. 169(1), pp. 131–138.

IRELAND, P.N. (2004): A method for taking models to the data. *Journal of Economic Dynamics and Control*, Vol. 28, pp. 1205–1226.

JANG, T.-S. AND SACTH, S (2012): *Identification of animal spirits in a bounded rationality model: an application to the Euro area*. MPRA Paper No. 37399, University Library of Munich, Germany.

LINDÉ, J. (2005): Estimating New-Keynesian Phillips curves: a full information maximum likelihood approach. *Journal of Monetary Economics*, Vol. 52, pp. 1135–49.

LINHART, H. AND ZUCCHINI, W. (1986): *Model Selection*. New York: John Wiley & Sons Inc.

MARMER, V. AND OTSU, T. (2012): Optimal comparison of misspecified moment restriction models under a chosen measure of fit. *Journal of Econometrics*, Vol. 170(2), pp. 538–550.

NASON, J.M. AND SMITH, G.W. (2008): Identifying the New Keynesian Phillips curve. *Journal of Applied Econometrics*, Vol. 23, pp. 525–551.

NEWKEY, W. AND WEST, K. (1987), A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, Vol. 55, pp. 703–708.

RABANALA, P. AND RUBIO-RAMIREZ, J.F. (2005): Comparing New Keynesian models of the business cycle: a Bayesian approach. *Journal of Monetary Economics*, Vol. 52(6), pp. 1151–1166.

RIVERS, D. AND VUONG, Q. (2002): Model selection tests for nonlinear dynamic models. *Econometrics Journal*, Vol. 5, pp. 1–39.

ROTEMBERG, J. AND WOODFORD, M. (1997): An optimization-based econometric framework for the evaluation of monetary policy. *NBER Macroeconomics Annual*, Cambridge, MA: MIT Press.

RUDD, J. AND WHELAN, K. (2005): New tests of the New-Keynesian Phillips curve. *Journal of Monetary Economics*, Vol. 52, pp. 1167–1181.

RUDD, J. AND WHELAN, K. (2006): Can rational expectations sticky-price models explain inflation dynamics. *American Economic Review*, Vol. 96(1), pp. 303–320.

SMETS, F. AND WOUTERS, R. (2003): An estimated stochastic dynamic general equilibrium model of the Euro area. *Journal of European Economic Association*, Vol. 1(5), pp. 1123–1175.

SMETS, F. AND WOUTERS, R. (2005): Comparing shocks and frictions in US and Euro area business cycles: a Bayesian DSGE approach. *Journal of Applied Econometrics*, Vol. 20(2), pp. 161–183.

SMETS, F. AND WOUTERS, R. (2007): Shocks and frictions in US business cycles: a Bayesian DSGE approach. *American Economic Review*, Vol. 97(3), pp. 586–606.

VUONG, Q. (1989): Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, Vol. 57(2), pp. 307–333.

WOODFORD, M. (2003): *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, New Jersey: Princeton University Press.

Reference to Chapter 4

AHRENS, S. AND SACHT, S. (2012): *Estimating a High-Frequency New Keynesian Phillips Curve*, Kiel Working Paper 1686, Kiel Institute for the World Economy, March 2012.

AKERLOF, G. AND SHILLER, R. (2009): *Animal Spirits. How Human Psychology Drives the Economy and Why it Matters for Global Capitalism*, New Jersey: Princeton University Press.

ALTONJI, J. AND SEGAL, L. (1996): Small-sample Bias in GMM Estimation of Covariance Structures. *Journal of Business and Economic Statistics*, Vol. 14, pp. 353–366.

ANATOLYEV, S. AND GOSPODINOV, N. (2011): *Methods for Estimation and Inference in Modern Econometrics*. New York: A Chapman and Hall Book.

ANDREWS, D. (1991): Heteroscedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, Vol. 59, pp. 817–858.

ASCARI, G. AND ROPELE, T. (2009): Trend Inflation, Taylor Principle, and Indeterminacy, *Journal of Money, Credit and Banking*, Vol. 41 (8), pp. 1557–1584.

ASSENMACHER-WESCHE, K. AND GERLACH, S. (2008): Interpreting Euro Area Inflation at High and Low Frequencies, *European Economic Review*, Vol. 52, pp. 964–986.

BINDER, M. AND PESARAN, M.H. (1995): Multivariate rational expectations models and macroeconomic modelling: A review and some new results. In: Pesaran, M.H. and Wickens, M. (ed.), *Handbook of Applied Econometrics: Macroeconomics*, Oxford: Basil Blackwell, pp. 139–187.

BRANCH, W.A. AND MCGOUGH, B. (2009): A New Keynesian model with heterogeneous expectations, *Journal of Economic Dynamics & Control*, Vol. 33, pp. 1036–1051.

CAMERER, C. (1998): Bounded Rationality in Individual Decision Making, *Experimental Economics*, Vol. 1, pp. 166–183.

CASTELNUOVO, E. (2010): Trend Inflation and Macroeconomic Volatilities in the Post WWII U.S. Economy, *North American Journal of Economics and Finance*, Vol. 21, pp. 19–33.

CHIARELLA, C. AND HE, T. (2002): Heterogeneous Beliefs, Risk and Learning in

a Simple Asset Pricing Model. *Computational Economics*, Vol. 19, pp. 95–132.

CHRISTIANO, L., EICHENBAUM, M. AND EVANS C. (2005): Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy*, Vol. 113(1), pp. 1–45.

COGLEY, T. AND SBORDONNE, A.M. (2008): Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve, *The American Economic Review*, Vol. 98 (5), pp. 2101–2126.

COGLEY, T. PRIMICERI, G. E. AND SARGENT, T. J. (2010), Inflation-Gap Persistence in the US, *American Economic Journal: Macroeconomics*, Vol. 2(1), pp. 43–69.

DAVIDSON, R. AND MACKINNON, G. (2004): *Econometric Theory and Methods*, New York: Oxford University Press.

DE GRAUWE, P. (2011): Animal Spirits and Monetary Policy , *Economic Theory*, Vol. 47(2–3), pp. 423–457.

DEJONG, D.N. AND DAVE, C. (2007): *Structural Macroeconomics*, New Jersey: Princeton University Press.

DUFFIE, D. AND SINGLETON, K. (1993): Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica*, Vol. 61(4), pp. 929–952.

DEL NEGRO, M. AND SCHORFHEIDE, F. (2008): Forming priors for DSGE models (and how it affects the assessment of nominal rigidities), *Journal of Monetary Economics*, Vol. 55, pp. 1191–1208.

EVANS, G. AND HONKAPOHJA, S. (2001): *Learning and Expectations in Macroeconomics*, New Jersey: Princeton University Press.

FAGAN, G., HENRY, J. AND MESTRE, R. (2001): *An Area Wide Model (AWM) for the Euro Area*, ECB working paper 42.

FORSELLS, M. AND KENNY, G. (2004): Survey Expectations, Rationality and the Dynamics of Euro Area Inflation, *Journal of Business Cycle Measurement and Analysis*, Vol. 2004(1), pp. 13–41.

FRANKE, R. (forthcoming): Microfounded Animal Spirits in the New Macroeconomic Consensus, *Studies in Nonlinear Dynamics and Econometrics*, forthcoming.

FRANKE, R., JANG, T.-S. AND SACHT, S. (2011): *Moment Matching versus Bayesian Estimation: Backward-Looking Behaviour in the New-Keynesian Three-*

Equations Model, Economics Working Paper 2011–10, Department of Economics, Christian-Albrechts-University of Kiel, Oktober 2011.

GOLDBAUM, D. AND MIZRACH, B. (2008): Estimating the Intensity of Choice in a Dynamic Mutual Fund Allocation Decision. *Journal of Economic Dynamics and Control*, Vol. 32(11), pp. 3866–3876.

HOMMES, H. (2006): Heterogeneous Agent Models in Economics and Finance, in: Tesfatsion, L. and Judd, K.(ed.), *Handbook of Computational Economics*, Vol. 2, Chapter. 23, pp. 1109-1186.

HOMMES, C. (2011): The Heterogeneous Expectations Hypothesis: Some Evidence from the Lab, *Journal of Economic Dynamics and Control*, Vol. 35 (1), pp. 1–24.

JANG, T.-S. (2012): Structural Estimation of the New-Keynesian Model: a Formal Test of Backward- and Forward-looking Behavior, in: N. Balke, F. Canova, F. Milani, and M. Wynne (ed.), *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Development*, *Advances in Econometrics: Vol. 28*, Emerald Group Publishing Limited, pp. 421–467.

KAHNEMAN, D. (2003): Maps of Bounded Rationality: Psychology for Behavioral Economics, *The American Economic Review*, Vol. 93(5), pp. 1449–1475.

KEYNES, J. M. (1936): *The General Theory of Employment, Interest and Money*, Macmillan Cambridge University Press.

LEE, B. S. AND INGRAM, B. (1991): Simulation estimation of time-series models, *Journal of Econometrics*, Vol. 47(2-3), pp. 197–205.

LEGNICK, M. AND WOHLTMANN, H.-W. (forthcoming): Agent-based Financial Markets and New Keynesian Macroeconomics - A Synthesis, *Journal of Economic Interaction and Coordination*, forthcoming.

LOMBARDI, M.J. AND NICOLETTI, G. (2011): Bayesian prior elicitation in DSGE models: Macro- vs micropriors, *Journal of Economic Dynamics & Control*, Vol. 36(2), pp. 294–313.

LUX, T. (2009): Rational Forecasts or Social Opinion Dynamics? Identification of Interaction Effects in a Business Climate Survey, *Journal of Economic Behavior and Organization*, Vol. 72(2), pp. 638–655.

MANKIW, N.G. AND REIS, R. (2002): Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve, *The Quarterly Journal of Economics*, Vol. 117(4), pp. 1295–1328.

MOONS, C., GARRETSEN, H., VAN AARLE, B. AND FORNERO, J. (2007): Monetary policy in the New-Keynesian model: An application to the Euro Area, *Journal of Policy Modeling*, Vol. 29(6), pp. 879–902.

NEWKEY, W. AND WEST, K. (1987): A Simple Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, Vol. 55, pp. 703–708.

NEWKEY, W. AND WEST, K. (1994): Automatic Lag Selection in Covariance Matrix Estimation, *Review of Economic Studies*, Vol. 61(4), pp. 631–53.

ROOS, M. W. M., SCHMIDT, U. (2012): The Importance of Time-Series Extrapolation for Macroeconomic Expectations, *German Economics Review*, Vol. 13(2), pp. 196–210.

RUSSELL, B. AND BANERJEE, A. (2008): The Long-run Phillips Curve and Non-stationary Inflation, *Journal of Macroeconomics*, Vol. 30, pp.1792-1815.

SARGENT, T.J. (1994): *Bounded Rationality in Macroeconomics*, New York: Oxford.

SELTEN, R. (2001): What is Bounded Rationality?, in: GIGERENZER, G. AND SELTEN, R. (Ed.) (2001): *Bounded rationality: The adaptive toolbox*, Cambridge: The MIT Press, pp. 13–36.

SIMS, C. (2003): Implications of Rational Inattention, *Journal of Monetary Economics*, Vol. 50(3), pp. 665–690.

SMETS, F. AND WOUTERS, R. (2003): An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area, *Journal of the European Economic Association*, Vol. 1, pp. 1123–1175.

SMETS, F. AND WOUTERS, R. (2005): Comparing Shocks and Frictions in US and Euro Area Business Cycles: a Bayesian DSGE Approach, *Journal of Applied Econometrics*, Vol. 20(2), pp. 161–183.

SMETS, F. AND WOUTERS, R. (2007): Shocks and frictions in US business cycles: A Bayesian DSGE approach, *American Economic Review*, Vol. 97, pp. 586–606.

SMITH, A. A. (1993): Estimating nonlinear time-series models using simulated vector autoregressions, *Journal of Applied Econometrics*, Vol. 18(S1), pp. 63–84.

WESTERHOFF, F. (2008): The Use of Agent-Based Financial Market Models to

Test the Effectiveness of Regulatory Policies, *Journal of Economics and Statistics*,
Vol. 228, pp. 195-227.

Reference to Chapter 5

AMILON, H. (2008): Estimation of an Adaptive Stock Market Model with Heterogeneous Agents. *Journal of Empirical Finance*, Vol. 15, pp. 342–362.

ANDERSON, S. AND PALMA, A. AND THISSE, J. (1992): *Discrete Choice Theory of Product Differentiation*. Cambridge, London: The MIT Press.

BROCK, W. AND HOMMES, C. (1997): Rational Routes to Randomness. *Econometrica*, Vol. 65, pp. 1059–1095.

BROCK, W. AND HOMMES, C. (1998): Heterogeneous Beliefs and Routes to Chaos in a Simple Asset Pricing Model. *Journal of Economic Dynamics and Control*, Vol. 22, pp. 1235–1274.

BÜHLMANN, P. (2002): Bootstraps for Time Series. *Statistical Science*, Vol. 17(1), pp. 52–72.

CHIARELLA, C. AND HE, X. (2002): Heterogeneous Beliefs, Risk and Learning in a Simple Asset Pricing Model. *Computational Economics*, Vol. 19, pp. 95–132.

DE GRAUWE, P. AND GRIMALDI, M. (2006): Exchange Rate Puzzles: A Tale of Switching Attractors. *European Economic Review*, Vol. 50, pp. 95–132.

DE LONG, J. AND SHLEIFER, A. AND SUMMERS, L. AND WALDMANN, R. (1990): Noise Traders Risk in Financial Markets. *Journal of Political Economy*, Vol. 98, pp. 703–738.

DUFFIE, D. AND SINGLETON, K. (1993): Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, Vol. 61(4), pp. 929–52.

FAGIOLO, G. AND BIRCHENHALL, C. AND WINDRUM, P. (2007): Special Issue on "Empirical Validation in Agent-Based Models". Eds., *Computational Economics*, Vol. 30(3).

FRANKE, R. (2010): On the Specification of Noise in Two Agent-based Asset Pricing Models. *Journal of Economic Dynamics and Control*, Vol. 34(6), pp. 1140–1152.

FRANKE, R. AND WESTERHOFF, F. (2011a): Estimation of a Structural Volatility Model of Asset Pricing. *Computational Economics*, Vol. 38, pp. 53–83.

FRANKE, R. AND WESTERHOFF, F. (2011b): *Structural Volatility in Asset Pricing Dynamics: Estimation and Model Contest*. mimeo, University of Kiel.

-
- FRANKE, R., AND WESTERHOFF, F. (2012): Converse Trading Strategies, Intrinsic Noise and the Stylized Facts of Financial Markets. *Quantitative Finance*, Vol. 12(3), pp. 425–436.
- GAUNERSDORFER, A. (2000): Endogenous Fluctuations in a Simple Asset Pricing Model with Heterogeneous Agents. *Journal of Economic Dynamics and Control*, Vol. 24, pp. 799–831.
- GAUNERSDORFER, A. AND HOMMES, C. (2007): A Nonlinear Structural Model for Volatility Clustering. In: Teyssiere, G. and Kirman, A. (ed.), *Long Memory in Economics*, Springer, pp. 265–288.
- GAUNERSDORFER, A. AND HOMMES, H. AND WAGENER, F. (2008): Bifurcation Routes to Volatility Clustering under Evolutionary Learning. *Journal of Economic Behavior & Organization*, Vol. 67(1), pp. 27–47.
- GILLI, M. AND WINKER, P. (2003): A Global Optimization Heuristic for Estimating Agent Based Models. *Computational Statistics and Data Analysis*, Vol. 42(3), pp. 299–312.
- GOLDBAUM, D. AND MIZRACH, B. (2008): Estimating the Intensity of Choice in a Dynamic Mutual Fund Allocation Decision. *Journal of Economic Dynamics and Control*, Vol. 32(11), pp. 3866–3876.
- HNATKOVSKA, V. AND MARMER, V. AND TANG, Y. (2011): *Supplement to "Comparison of misspecified calibrated models"*. Working paper, University of British Columbia.
- HNATKOVSKA, V. AND MARMER, V. AND TANG, Y. (2012): Comparison of Misspecified Calibrated Models: the Minimum Distance Approach. *Journal of Econometrics*, *Forthcoming*.
- LAHIRI, S.N. (2003): *Resampling Methods for Dependent Data*. Berlin: Springer.
- LEE, B.-S. AND INGRAM, B.F. (1991): Simulation Estimation of Time-series Models. *Journal of Econometrics*, Vol. 47, pp. 197–205.
- LEE, S. AND LAI, P. (2009): Double Block Bootstrap Confidence Intervals for Dependent Data. *Biometrika*, Vol. 96(2), pp. 427–443.
- LONGTIN, A. (2003): *Effect of noise on nonlinear dynamics*, A. Beuter, L. Glass, M.C. Mackey, M. Titcombe (Eds.), *Nonlinear Dynamics in Physiology and Medicine*, Springer-Verlag, New York, pp. 149–189.

MANSKI, C. AND MCFADDEN, D. (1981): *Structural Analysis of Discrete Data and Econometric Applications*. Cambridge: The MIT Press.

PAGAN, A. (1996): The Econometrics of Financial Markets. *Journal of Empirical Finance*, Vol. 3, pp. 15–102.

VUONG, Q. (1989): Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses. *Econometrica*, Vol. 57(2), pp. 307–333.

WINKER, P. AND GILLI M. AND JELESKOVIC V. (2007): An Objective Function for Simulation Based Inference on Exchange Rate Data. *Journal of Economic Interaction and Coordination*, Vol. 2, pp. 125–145.

WINKER, P. AND JELESKOVIC, V. (2006): *The Unconditional Distribution of Exchange Rate Returns: Statistics, Robustness, Time Aggregation*. Working paper 008-06, University of Essex, Colchester.

WINKER, P. AND JELESKOVIC, V. (2007): *Dependence of - and Long Memory in - Exchange Rate Returns: Statistics, Robustness, Time Aggregation*. Working paper 011-07, University of Essex, Colchester.

Reference to Chapter 6

ALFARANO, S. AND LUX, T. (2007): A Noise Trader Model as a Generator of Apparent Financial Power Laws and Long Memory, *Macroeconomic Dynamics*. Vol. 11, pp. 80–101.

ALFARANO, S., AND LUX, T. AND WAGNER, F. (2005): Estimation of Agent-Based Models: The Case of an Asymmetric Herding Model, *Computational Economics*. Vol. 26, pp. 19–49.

ALFARANO, S., AND LUX, T. AND WAGNER, F. (2007): Empirical Validation of Stochastic Models of Interacting Agent, *The European Physical Journal B*. Vol. 55, pp. 183–187.

ALFARANO, S., AND LUX, T. AND WAGENER, F. (2008): Time-Variation of Higher Moments in a Financial Market with Heterogeneous Agent: an Analytical Approach. *Journal of Economic Dynamics and Control*, Vol. 32(1), pp. 101–136.

AMILON, H. (2008): Estimation of an Adaptive Stock Market Model with Heterogeneous Agents. *Journal of Empirical Finance*, Vol. 15, pp. 342–362.

ANDERSEN, T. AND B. SØRENSEN (1996): GMM Estimation of a Stochastic Volatility Model: a Monte Carlo Study. *Journal of Business and Economic Statistics*, Vol. 14(3), pp. 328–352.

ANDREWS, D. (2004), The Block-Block Bootstrap: Improved Asymptotic Refinements. *Econometrica*, Vol. 72(3), pp. 673–700.

BROCK, W. AND HOMMES, C. (1997): Rational Routes to Randomness. *Econometrica*, Vol. 65, pp. 1059–1095.

BROCK, W. AND HOMMES, C. (1998): Heterogeneous Beliefs and Routes to Chaos in a Simple Asset Pricing Model. *Journal of Economic Dynamics and Control*, Vol. 22, pp. 1235–1274.

CARLSTEIN, E. (1986): The Use of Subseries Methods for Estimating the Variance of a General Statistics from a Stationary Time Series. *The Annals of Statistics*, Vol. 14, pp. 1171–1179.

CARRASCO, M. AND FLORENS, J. (2002): Simulation-Based Method of Moments and Efficiency. *Journal of Business and Economic Statistics*, Vol. 20(4), pp. 482–492.

CHEN, S.-H, AND CHANG, C.-L, AND DU, Y.-R. (2010): Agent-Based Economic Models and Econometrics. *Knowledge Engineering Review*.

- CHIARELLA, C. AND HE, X. (2002): Heterogeneous Beliefs, Risk and Learning in a Simple Asset Pricing Model. *Computational Economics*, Vol. 19, pp. 95–132.
- CHIARELLA, C., AND HE, X. AND WANG, D. (2006): A Behavioral Asset Pricing Model with a Time-Varying Second Moment. *Chaos, Solitons and Fractals*, Vol. 29, pp. 535–555.
- DAVIDSON, R. AND MACKINNON, J.G. (2004): *Econometric Theory and Methods*. Oxford: Oxford University Press.
- DAY, R., AND HUANG, W. (1990): Bulls, Bears and Market Sheep. *Journal of Economic Behavior & Organization*, Vol. 14(3), pp. 299–329.
- DE GRAUWE, P., AND H. DEWACHTER, AND M. EMBRECHTS (1993): *Exchange Rate Theory: Chaotic Models of Foreign Exchange Markets*. Blackwell Publishers.
- DE GRAUWE, P. AND GRIMALDI, M. (2006): Exchange Rate Puzzles: A Tale of Switching Attractors. *European Economic Review*, Vol. 50, pp. 95–132.
- DE LONG, J. AND SHLEIFER, A. AND SUMMERS, L. AND WALDMANN, R. (1990): Noise Traders Risk in Financial Markets. *Journal of Political Economy*, Vol. 98, pp. 703–738.
- DUFFIE, D. AND SINGLETON, K. (1993): Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, Vol. 61(4), pp. 929–52.
- FÖLLMER, H., AND U. HORST, AND A. KIRMAN (2005): Equilibria in Financial Markets with Heterogeneous Agents: a Probabilistic Perspective. *Journal of Mathematical Economics*, Vol. 41, pp. 123–155.
- FRANKE, R., AND WESTERHOFF, F. (2009): *Converse Trading Strategies, Intrinsic Noise and the Stylized Facts of Financial Markets*. Working paper, University of Kiel and University of Bamberg.
- FRANKE, R. AND WESTERHOFF, F. (2011): Estimation of a Structural Volatility Model of Asset Pricing. *Computational Economics*, Vol. 38, pp. 53–83.
- FRANKEL, J.A. AND FROOT, K.A. (1990): The Rationality of the Foreign Exchange Rate: Chartists, Fundamentalists and Trading in the Foreign Exchange Market. *American Economic Review*, Vol. 80(2), pp. 181–185.
- GAUNERSDORFER, A. (2000): Endogenous Fluctuations in a Simple Asset Pricing Model with Heterogeneous Agents. *Journal of Economic Dynamics and Control*,

Vol. 24, pp. 799–831.

GAUNERSDORFER, A. AND HOMMES, C. (2007): A Nonlinear Structural Model for Volatility Clustering. In: Teyssiere, G. and Kirman, A. (ed.), *Long Memory in Economics*, Springer, pp. 265–288.

GAUNERSDORFER, A. AND HOMMES, H. AND WAGENER, F. (2008): Bifurcation Routes to Volatility Clustering under Evolutionary Learning. *Journal of Economic Behavior & Organization*, Vol. 67(1), pp. 27–47.

GEWEKE, J. AND PORTER-HUDAK, S. (1983): The Estimation and Application of Long-memory Time Series Models. *Journal of Time Series Analysis*, Vol. 4, pp. 221–237.

GILLI, M. AND WINKER, P. (2003): A Global Optimization Heuristic for Estimating Agent Based Models. *Computational Statistics and Data Analysis*, Vol. 42(3), pp. 299–312.

HALL, P. AND HOROWITZ, J. (1996): Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators. *Econometrica*, Vol. 64(4), pp. 891–916.

HILL, B (1975): A Simple Approach to Inference about the Tail of a Distribution. *Annals of Statistics*, Vol. 3, pp. 1163–1174.

HOMMES, C. (2001): Financial Markets as Nonlinear Adaptive Evolutionary Systems. *Quantitative Finance*, Vol. 1, pp. 149–167.

HOMMES, C. (2006): Heterogeneous Agent Models in Economics and Finance. In: Tesfason, L. and Judd K.L. (ed.), *Handbook of Computational Economics*, Vol. 2, Chapter. 23.

HORST, U. AND ROTHE, C. (2008): Queuing, Social Interactions, And The Microstructure Of Financial Markets. *Macroeconomic Dynamics*, Vol. 12, pp. 211–233.

INOUE, A. AND SHINTANI, M. (2006): Bootstrapping GMM Estimators for Time Series. *Journal of Econometrics*, Vol. 133(2), pp. 531–555.

JONSSON, G. AND KLEIN, P. (1996): Stochastic Fiscal Policy and the Swedish Business Cycle. *Journal of Monetary Economics*, Vol. 38, pp. 245–268.

KEYNES, J.M., *The General Theory of Employment, Interest and Money*. Macmillan Cambridge University Press.

KIRMAN, A. (1991): Epidemics of Opinion and Speculative Bubbles in Financial Markets. In: Taylor, M.(ed.), *Money and Financial Markets*, New York: Macmillan.

KIRMAN, A. (1993): Ants, Rationality, and Recruitment. *Quarterly Journal of Economics*, Vol. 108, pp. 137–156.

LUX, T. (1995): Herd Behavior, Bubbles and Crashes. *Economic Journal*, Vol. 105, pp. 881-896.

LUX, T. (1997): Time Variation of Second Moments from a Noise Trader/Infection Model. *Journal of Economic Dynamics and Control*, Vol. 22, pp. 1–38.

LUX, T. (2006): Financial Power Laws : Empirical Evidence, Models, and Mechanisms. In: Cioffi, C. (ed.), *Power Laws in the Social Sciences: Discovering Complexity and Non-Equilibrium Dynamics in the Social Universe*, Cambridge University Press.

LUX, T. (2009): Stochastic Behavioral Asset Pricing Models and the Stylized Facts. In: Thorsten Hens, Klaus Reiner Schenk-Hoppe (ed.), *Handbook of Financial Markets: Dynamics and Evolution*, Amsterdam: North-Holland Pub. Co.

MICHAELIDES, A. AND NG, S. (2000): Estimating the Rational Expectations Model of Speculative Storage: a Monte Carlo Comparison of Three Simulation Estimators. *Journal of Econometrics*, Vol. 96, pp. 231–266.

NEWKEY, W. AND WEST, K. (1987), A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, Vol. 55, pp. 703–708.

NEWKEY, W. AND WEST, K. (1994): Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, Vol. 61(4), pp. 631–53.

POLLARD, M. (2006): *Stat 251 Paper Review* on "A Minimal Noise Trader Model as a Generator of Apparent Financial Power Laws and Long Memory", <http://www.matthewcpollard.com>.

WESTERHOFF, F. (2009): *A Simple Agent-based Financial Market Model: Direct Interactions and Comparisons of Trading Profits*. BERG Working Paper Series No. 61, University of Bamberg.

WINKER, P. AND JELESKOVIC, V. (2006): *The Unconditional Distribution of Exchange Rate Returns: Statistics, Robustness, Time Aggregation*. Working paper 008-06, University of Essex, Colchester.

WINKER, P. AND JELESKOVIC, V. (2007): *Dependence of - and Long Memory in - Exchange Rate Returns: Statistics, Robustness, Time Aggregation*. Working paper 011-07, University of Essex, Colchester.

WINKER, P. AND GILLI M. AND JELESKOVIC V. (2007): An Objective Function for Simulation Based Inference on Exchange Rate Data. *Journal of Economic Interaction and Coordination*, Vol. 2, pp. 125–145.

WINKER, P. AND MARINGER, D. (2006): The Threshold Accepting Optimization Algorithm in Economics and Statistics. In: Kontoghiorges, E. and Gatu, C. (Hg.), *Advances in Computational Economics, Finance and Management Science*.

ZEEMAN, E. (1974): On the Unstable Behaviour of Stock Exchanges. *Journal of Mathematical Economics*, Vol. 1(1), pp. 39–49.

Reference to Chapter 7

ALFARANO, S. AND LUX, T. (2007): A Noise Trader Model as a Generator of Apparent Financial Power Laws and Long Memory, *Macroeconomic Dynamics*. Vol. 11, pp. 80–101.

DE GRAUWE, P. (2011): Animal spirits and monetary policy. *Economic Theory*, Vol. 47, pp. 423–457.

FRANKE, R., AND WESTERHOFF, F. (2012): Converse Trading Strategies, Intrinsic Noise and the Stylized Facts of Financial Markets. *Quantitative Finance*, Vol. 12(3), pp. 425–436.

GAUNERSDORFER, A. AND HOMMES, C. (2007): A Nonlinear Structural Model for Volatility Clustering. In: Teyssiere, G. and Kirman, A. (ed.), *Long Memory in Economics*, Springer, pp. 265–288.

SMETS, F. AND WOUTERS, R. (2003): An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area, *Journal of the European Economic Association*, Vol. 1, pp. 1123–1175.

SMETS, F. AND WOUTERS, R. (2005): Comparing Shocks and Frictions in US and Euro Area Business Cycles: a Bayesian DSGE Approach, *Journal of Applied Econometrics*, Vol. 20(2), pp. 161–183.

Curriculum Vitae

Tae-Seok Jang

Education

2012: Ph.D in Economics, University of Kiel, Germany
2008/2009: Advanced Studies Program, Kiel Institute for the World Economy, Germany
2007: Quantitative Economics Program, University of Kiel, Germany
2006: M.A, International Area Studies (Economics), GSIS, Seoul National University, Korea
2004: B.A, German, Educational Department, Seoul National University, Korea
1997: High School Diploma, Myung-Duk Foreign Language, Korea

Research Interest

International Finance, Computational Economics, Estimation of Agent-based Models, Applied Macroeconomics (Business cycle)

Awards

2012: One of three best papers in the Graduate Student Contest of the Society of Computational Economics 2012
2011: Möller fund for conference stay, University of Kiel (CAU)
2007-2011: German Academic Exchange Service Research Grant (DAAD)
2005: Prize at thesis contest of Korea International Trade Association (KITA)
2003-2006: University scholarship, Seoul National University

Publication

Refereed Work

- Jang, T.-S. (2012): Structural Estimation of the New-Keynesian Model: a Formal Test of Backward- and Forward-looking Behavior, in: N. Balke, F. Canova, F. Milani, and M. Wynne (ed.), *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Development*, *Advances in Econometrics: Vol. 28*, Emerald Group Publishing Limited, 421–467.
- Franke, R., T.-S. Jang, & S. Sacht (2011), *Moment Matching versus Bayesian Estimation: Backward-looking Behaviour in the New-Keynesian Three-Equations Model*. Economic Working paper 2011-10, University of Kiel [preparing a submission].
- Jang, T.-S. & S. Sacht (2011), *Identification of Animal Spirits in a Bounded Rationality Model: An Application to the Euro Area*. MPRA Paper No. 37399, University Library of Munich, Germany [preparing a submission].
- Sato, A.-H., J. Ghonghadze, T.-S. Jang, F. Wagner, & T. Lux (2010), *Parameter Estimation Methods of a Multiplicative Stochastic Process for Financial Time Series Analysis: an Application to Inference of Tail-Risks*, mimeo, University of Kiel [preparing a submission].

Non-refereed work

- Ambrocio, G., & T.-S. Jang (2009), Productivity shocks and New Keynesian Phillips curve: evidence from US and Euro area, Kiel Advanced Studies Working Paper Series No.453
- Jang, T.-S. (2008), *Makro-ökonomische Modelle: Komplexe Wirtschaftssysteme und Soziale Interaktion*, Japanisch-Deutsches Zentrum Berlin, Band 58
- Jang, T.-S. (2007) South Korea's aid to North Korea's transformation process, *Discourses in Social Market Economy* 2007-06

Professional Experience

2011.4 - 2012.3: Research Assistant in University of Kiel (CAU)

2006.9 - 2007.5: Research Assistant in Korea Development Institute (KDI)

2005.3 - 2006.8: Research Assistant in Seoul National University (SNU)

Teaching Experience

Tutorial in International Monetary Economics Summer term 2011 (CAU)

Tutorial in Theory of Financial Markets Winter term 2011/12 (CAU)

Invited Talks

- Year 2012

- Seminar at the Bank of Korea

- The SCE 18th International Conference on Computing in Economics and Finance, Prague, Czech Republic

- Statistical/Econometric Seminar, University of Kiel

- Year 2011

- 5th CSDA International Conference on Computational and Financial Econometrics (CFE'11), Senate House, University of London, UK

- Macroeconometric Workshop, German Institute for Economic Research (DIW), Berlin

- Statistical/Econometric Seminar, University of Kiel

- 10th Annual Advances in Econometrics Conference at SMU "DSGE Models in Macroeconomics: Estimation, Evaluation and New Developments", USA

- Year 2009 and 2010

- ZEW Summer Workshop on Expectations and Expectation Formation in Financial Markets, Mannheim, Germany

- Statistical/Econometric Seminar, University of Kiel

- Year 2007 and 2008

- Advanced Studies Program Conference, Kiel Institute for the World Economy

- German-Japanese-Korea Grantee Seminar, Berlin

- Ordnungspolitischer Dialog: Das Konzept der Sozialen Marktwirtschaft und seine Anwendung, Zwickau

Languages

Written and spoken English (Fluent), German (Advanced), Korean (Mother Tongue)

Computer experience

Experience with R, E-Views, Stata, Matlab, GAUSS, C/C++, Fortran 95, Shell programming, parallel computation using message passing interface (MPI) library

Eidesstattliche Erklärung

Ich erkläre hiermit, an Eides Statt, dass ich meine Doktorarbeit "Moment-Based Estimation of Macroscopic Dynamic Models in Macroeconomics and Finance" selbständig und ohne fremde Hilfe angefertigt habe und dass ich alle von anderen Autoren wörtlich übernommenen Stellen, wie auch die sich an die Gedanken anderer Autoren eng anlehenden Ausführungen meiner Arbeit, besonders gekennzeichnet und die Quellen nach den mir angegebenen Richtlinien zitiert habe.

Kiel, 28 August 2012

Tae-Seok Jang

