CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Monetary Policy Under Behavioral Expectations: An Empirical Validation of the Heuristic Switching Model

Bachelor's Thesis

Author: Sergey Bolshakov
Study program: Economics and Finance
Supervisor: Jiří Kukačka, Ph.D.
Year of defense: 2020

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, May 7, 2020

Sergey Bolshakov

Abstract

This work takes one of the most prominent behavioral New-Keynesian models from the shelf and estimates it via the simulated method of moments. The model exhibits a remarkably good fit to the auto- and cross-covariance profiles of the euro area macroeconomic time series, especially compared to the standard rational expectations model. This result corroborates the claim that central banks which implement strict inflation targeting are better off reacting to the output gap, on top of inflation.

JEL Classification	E52, E70, D84, C53
Keywords	Behavioral macroeconomics, heterogeneous expectations, New-Keynesian model, Heuristic Switching Model, simulated method of moments
Title	Monetary Policy Under Behavioral Expectations: An Empirical Validation of the Heuristic Switching Model

Abstrakt

Tato práce se zabývá odhadem jednoho z nejvýznačnějších behaviorálních keynesiánských modelů pomocí metody simulačńch momentů. Behaviorální model fituje kovarianční matici makroekonomických časových řad euro oblasti pozoruhodně dobře, zvlášť v porovnání se standardním racionálním modelem. Tento výsledek podporuje tvrzení, že centrální banky, které se pečují o cenovou stabilitu, mohou dosáhnout tohoto cíle lépe, pokud budou reagovat nejen na inflaci, ale i na mezeru výstupu.

Klasifikace JEL	E52, E70, D84, C53
Klíčová slova	Behaviorální makroekonomie, heterogenní očekávání, nový keynesiánský model, heuristický model, metoda simulačních momentů
Název práce	Monetární politika za předpokladu behaviorálních očekávání: Empirická validace heurestického modelu

Acknowledgments

The author is grateful especially to Jiří Kukačka, Ph.D., for his wise guidance, incredible support and painstaking reviews of the drafts. Special thanks for stimulating discussions go to Priv.-Doz. Dr. Reiner Franke. Assistance with implementation issues provided by Domenico Massaro, Ph.D., is also appreciated. None of the above are responsible for errors in this work.

Typeset in LaTeXusing the IES Thesis Template.

Bibliographic Record
Bolshakov, Sergey: *Monetary Policy Under Behavioral Expectations: An Empirical Validation of the Heuristic Switching Model.* Bachelor's Thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2020, pages 41. Advisor: Jiří Kukačka, Ph.D.

Contents

List of Tables

List of Figures

Abbreviations

DGP data-generating process

DSGE dynamic stochastic general equilibrium

ECB European Central Bank

HSM Heuristic Switching Model

LtFE learning-to-forecast experiments

NKM New-Keynesian model

NKPC New-Keynesian Phillips curve

RE rational expectations

SMLE simulated maximum likelihood estimation

SMM simulated method of moments

Bachelor's Thesis Proposal

Author	Sergey Bolshakov
Supervisor	Jiří Kukačka, Ph.D.
Proposed topic	Monetary Policy Under Behavioral Expectations: An Empirical Validation of the Heuristic Switching Model

Research question and motivation Expectations play an important role in macroeconomics. Agent-based modeling provides an opportunity to incorporate virtually any expectations formation mechanisms into economic models and then analyze dynamic and stochastic properties of the models by running multiple simulations under different parameter settings.

One of such mechanisms is the 4-type Heuristic Switching Model (HSM) developed by Brock and Hommes (1997) and extended by Anufriev and Hommes (2012). Recently, Hommes et al. (2017) employed the HSM as the expectations formation mechanism in the baseline three-equation New Keynesian model to conduct monetary policy analysis and concluded that the current European Central Bank's (ECB) policy rule might be not the optimal one. The question this thesis will aim at examining is whether the model they used can be fitted to real macroeconomic data.

From a theoretical perspective, we would like to verify whether the experiments with human subjects which gave rise to the development of the HSM are suitable for inference about the expectations formation process in the real-world setting. In addition to that, the HSM has relatively many parameters, and so it will be interesting to examine the presence of a trade-off between accuracy and simplicity by comparing its empirical performance to the performance of the 'pure technical block' (PTB) from Jang and Sacht (2017), a simple 2-type expectations formation mechanism which outperformed all other models (including the hybrid version of the New Keynesian model) in their study in terms of fitness to the Euro area data.

Contribution From a practical perspective, this thesis will provide a robustness check for the monetary policy recommendations presented in Hommes et al. (2017). Because the authors mention the ECB as a potential beneficiary of their conclusions, it is important to examine whether the model used in their study fits macroeconomic data. The Euro area macroeconomic time series will be utilized, so the results of the thesis will be particularly applicable for the ECB policymakers.

From a theoretical perspective, the thesis can be viewed as an examination of the empirical performance of one of the most prominent expectations formation mechanisms in the literature on learning and bounded rationality. That is, the thesis will contribute to the field by shedding some light on the appropriateness of learning-to-forecast experiments with human subjects as a source of insights into individuals' expectations formation process in the context of macroeconomics. In addition, this thesis will effectively extend the work of Jang and Sacht (2017) by introducing a new scenario to their 'horse race'

Methodology The thesis will estimate the baseline three-equation New Keynesian model with the HSM as the expectations formation mechanism. As a benchmark to evaluate the model's empirical performance, the PTB will be employed. We will use the Area-Wide Model database (https://eabcn.org/page/area-wide-model) as the data source, and the Simulated Method of Moments as the estimation approach.

Outline

1. Introduction
2. Literature review
3. The model
4. Methodology
5. Data
6. Results and discussion
7. Conclusion

Bibliography

Anufriev, M. and Hommes, C. (2012): 'Evolutionary Selection of Individual Expectations and Aggregate Outcomes in Asset Pricing Experiments', *American Economic Journal: Microeconomics*, 4(4), pp. 35-64.

Brock, W. and Hommes, C. (1997): 'A Rational Route to Randomness', *Econometrica*, 65, pp. 1059-1095.

Fagan, G., Henry, J. and Mestre, R. (2005): 'An area-wide model for the Euro area', *Economic Modelling*, 22(1), pp. 39-59.

Franke, R., Jang, T.-S. and Sacht, S. (2015): 'Moment matching versus Bayesian estimation: backward-looking behavior in the New-Keynesian baseline model', *The North American Journal of Economics and Finance*, 31, pp. 126-154.

Hommes, C., Massaro, D. and Weber, M. (2017): 'Monetary Policy under Behavioral Expectations: Theory and Experiment', Bank of Lithuania Working Paper, No. 42/2017.

Jang, T.-S. and Sacht, S. (2017): 'Modeling consumer confidence and its role for expectation formation: A horse race', Kiel University Economics Working Paper, No. 2017-04.

Author

Supervisor

Chapter 1

Introduction

Expectations play an important role in macroeconomics. The New-Keynesian framework with the assumption of rational expectations (RE) has become the dominating paradigm in macroeconomic theory. Thanks to its analytical tractability, RE models provide a convenient framework for monetary policy analysis. However, this comes at a price, since the assumption that economy consists exclusively of cognitively unlimited agents with full understanding of the underlying model and perfect foresight about the future and other agents' decisions seems rather unrealistic.

One way to find a more solid expectation formation mechanism in this respect is to design an experiment in which the way the agents form their expectations about the future is directly observable. This is exactly what the so-called learning-to-forecast experiments (LtFE) are dealing with. A typical learning-to-forecast experiment consists of 6 human subjects whose goal is to predict the future value of a variable, most commonly the price of an asset. The participants are equipped with a basic grasp of how the underlying laboratory economy works and with a full knowledge of past realizations of the variable as well as his/her past predictions. Not only such experiments allow to gauge the true expectation formation process of the economy participants, but they also provide an opportunity to assess different monetary policy regimes without making any assumptions about the way agents form their expectations. The external validity of laboratory experiments may seem questionable at first glance, but, as argued by Cornand and Hubert (2020), the outcomes of learning-to-forecast experiments are comparable to other common sources of information about agents' expectations, such as surveys or financial market data. And this sort of reinforces the view of Hommes (2011) who concentrates on the LtFE literature and concludes that expectations of economic agents can hardly be classified as homogeneous and rational.

Aiming at explaining stylized facts from these laboratory experiments by one expectation formation mechanism, Anufriev and Hommes (2012) modify the discrete-choice model by Brock and Hommes (1997) and populate it with four heuristics, four rules-of-thumb which deliver the best fit to human subjects' forecasts. This set of heuristics includes both stabilizing and destabilizing rules which, when properly framed into the switching mechanism with some inertia in agents' decisions taken into account, are able to generate different patterns in time series. This is the main idea of the Heuristic Switching Model (HSM), one of the most prominent expectation formation mechanisms in the LtFE literature, both in asset-pricing experiments (Hommes et al. 2005; 2008), and – importantly enough – in the New-Keynesian macroeconomic environments (Assenza et al. 2019; Hommes et al. 2019).

It is no surprise, then, that Hommes et al. (2019) employ the HSM as the behavioral counterpart of RE in their optimal monetary policy exercise which reveals that central banks implementing strict inflation targeting are at risk or running a sub-optimal monetary policy, if they are assuming RE. More specifically, they find that, instead of reacting only to fluctuations in inflation, such central banks as the European Central Bank (ECB) are better off reacting to the output gap fluctuations on top of inflation, even though under the standard assumption of RE it might seem detrimental to its main priority, inflation stability. Although this claim is not entirely new in the literature (see De Grauwe 2011, among others), it is quite remarkable that the New-Keynesian model (NKM) in Hommes et al. (2019) is fully microfounded (another recent example is De Grauwe and Ji 2020). In addition to that, they provide a sort of robustness check to these results by conducting a learning-to-forecast experiment with human subjects where the only difference between the two treatments is the central bank's behavior with respect to the output gap fluctuations: reacting to the output gap in addition to inflation or not reacting at all. The authors find that all macroeconomic variables – the output gap, the inflation rate and the nominal interest rate – display more stable dynamics when the central bank reacts also to the output gap fluctuations.

However, it is important to note that the calibration of the NKM in Hommes et al. (2019) is taken from Clarida et al. (2000) who estimate the model on the US data, whereas their monetary policy recommendations are addressed to other central banks. Moreover, the parameters of the switching mechanism between the four heuristics are set in accordance with Anufriev and Hommes (2012) who arrive at this calibration “after some trial and error simulations”. And although the latter does not undermine the credibility of the LtFE part of their study (because in such experiments the expectations are formed directly by participants), together with the US calibration it may cast some doubts on the applicability of the results of the experimental as well as the theoretical exercise to the euro area, the territory where the ECB actually operates.

Even more importantly, a very basic question whether the NKM with HSM actually fits the euro area macroeconomic time series is still open. This is exactly the question this work aims at answering. As a natural benchmark, we make use of the standard RE model. And provided the HSM outperforms RE in terms of fitness to the macroeconomic data, we repeat the monetary policy exercise from Hommes et al. (2019) under the resulting ‘euro calibration’ to check the robustness of their optimal monetary policy recommendations with respect to the underlying calibration.

To execute this, we employ the simulated method of moments (SMM) along the lines of Jang and Sacht (2019; 2016) who are also concerned with estimation of a behavioral macroeconomic model. Simply put, SMM is a transparent and straightforward estimation technique which seeks to bring the model’s simulated moments as closely as possible to its empirical counterparts. As another important part of methodology, we utilize Monte Carlo simulations to verify that our estimation procedure is capable of recovering the pseudo-true parameters and has a sufficient power to discriminate between the two models.

In short, we find that HSM fits the empirical data much better than RE. The resulting ‘euro calibration’ yields qualitatively the same monetary policy implications as those in Hommes et al. (2019), namely, that central banks which target inflation should also react to the output gap fluctuations. This result seems especially relevant for the ECB. On the methodological side, a closer inspection of moment matching suggests that SMM may have some room for improvement, at least when dealing with behavioral macroeconomic models. As a preliminary response, three alternative fitness measures are proposed.

The contribution of this work is thus clear: essentially, it certifies that one of the most prominent behavioral macroeconomic models together with its main monetary policy implication is empirically valid. As a by-product, this work identifies a minor methodological caveat. To the best of our knowledge, these issues have not been addressed in the literature before. The most closely related work to ours, Jang and Sacht (2019), is concerned with behavioral specifications of NKM which are neither microfounded, nor have a solid support in macroeconomic LtFE literature, as opposed to the model investigated in this work.

The remainder of the work is organized as follows. The next chapter offers a brief overview of related literature. Chapter 3 introduces the behavioral macroeconomic model and demonstrates its monetary policy implications. Chapter 4 describes the estimation procedure and Monte Carlo simulations. Chapter 5 reports the results of empirical application and performs the robustness check of the implications for monetary policy. Chapter 6 provides a deeper assessment of the empirical results and discusses the methodological caveat. Chapter 7 concludes. Supporting materials for Chapters 3, 4, 5 are relegated to Appendices A, B and C, respectively. Online Appendix D contains the underlying scripts for Monte Carlo experiments, empirical estimation and reproduction of figures presented in this work.

Chapter 2

Literature Review

The history of estimation of the behavioral NKM is not long. We identify Milani (2007) as one of the main pioneers in this area. Being concerned with the well-known inability of NKM to reproduce the persistence observed in macroeconomic time series, he takes the hybrid NKM from the shelf and challenges its "mechanical" constructs, namely, the price indexation and habit formation terms. Using Bayesian approach and US data, he compares the empirical fitness of RE with that of constant-gain learning mechanism. He finds that, in presence of behavioral expectation formation, NKM does not need explicit backward-looking components to explain the inertia in empirical time series.

On the other hand, Liu and Minford (2014) find evidence that behavioral expectations do not really add value to the NKM framework. After arguing that Bayesian methods are not appropriate for testing different models against empirical data, they employ the method of indirect inference to compare the RE model with behavioral specification of NKM by De Grauwe (2010) on the US post-war data. They report that RE outperforms its behavioral counterpart decisively.

Meanwhile, having taken a taste of SMM in a simple asset-pricing model estimation exercise in Franke (2009), he, together with his younger colleagues, customizes this method for a hybrid NKM and contrasts it with Bayesian approach in Franke et al. (2015). They report that the two methods yield somewhat different results and seem to focus on different aspects of data and suggest that, thanks to its transparency, explicitness and lower computational burden, SMM is a more appropriate estimation method for such models as NKM. Building on the newly-developed moment-matching estimation framework, Jang and Sacht (2016) introduce behavioral expectations from De Grauwe (2011) into NKM and compare its empirical performance against the standard hybrid specification with backward-looking components. They find evidence for behavioral expectation formation in the euro area, as the corresponding model provides a solid fit to macroeconomic data. Jang and Sacht (2019) bring this aspiration to the next level by setting up a "horse race" of RE and different combinations of behavioral heuristics from De Grauwe (2011) and Gaunersdorfer et al. (2008) both on the euro area and US data. They conclude that expectations seem to be emotional for US and technical in nature for the euro area.

At the same time, Kukacka et al. (2018) bring the simulated maximum likelihood estimation (SMLE) to the multivariate macroeconomic framework by estimating behavioral model from De Grauwe (2011) and RE specification of NKM both on the US and euro area data. Their main contribution is that they manage to pin down the intensity of choice, one of the most challenging behavioral parameters to estimate. Focusing instead on one element of the New-Keynesian framework, the New-Keynesian Phillips curve (NKPC), Cornea-Madeira et al. (2019) provide a considerable evidence for behavioral heterogeneity in US inflation dynamics.

Finally, Franke (2019) demonstrates a more mature version of the SMM by estimating the Harrod-Kaldor business cycle model and effectively allows us to help materialize perhaps one of the biggest ambitions of Milani (2007) – to perform an empirical test of a truly microfounded behavioral version of NKM.

Chapter 3

The Model

This chapter describes the macroeconomic model of interest. The emphasis is put on the underlying expectation formation mechanism. The model's main implications for monetary policy are then examined. Basically, this chapter reiterates its counterpart from Hommes et al. (2019).

3.1 The Model

The slightly modified[1] version of the baseline 3-equation NKM from Hommes et al. (2019) reads as follows:

$$y_t = \bar{y}^e_{t+1} - \tau(r_t - \bar{\pi}^e_{t+1}) + \varepsilon_{y,t} \tag{3.1}$$

$$\pi_t = \nu\bar{\pi}^e_{t+1} + \kappa y_t + \varepsilon_{\pi,t} \tag{3.2}$$

$$r_t = \phi_\pi \pi_t + \phi_y y_t, \tag{3.3}$$

where y_t is the output gap, π_t is inflation rate gap and r_t is the nominal interest rate gap. Variables $\bar{y}^e_{t+1}$ and $\bar{\pi}^e_{t+1}$ are the expected output gap and inflation gap at time $t + 1$, respectively. Both represent averages over population of agents, and both are formed at time t when only information up to time $t - 1$ is available. Variables $\varepsilon_{y,t}$ and $\varepsilon_{\pi,t}$ are independent and identically distributed random shocks.

Equation (3.3) is the central bank's monetary policy rule, equation (3.2) is the NKPC, and equation (3.1) is the dynamic IS curve. ϕ_π and ϕ_y are non-negative parameters which determine how the central bank reacts to the shocks to inflation gap and the output gap, respectively. The shocks $\varepsilon_{y,t}$ and $\varepsilon_{\pi,t}$ are

[1] See Appendix A for description and discussion of the modifications made.

Table 3.1: The Set of Heuristics

1	ADA	Adaptive rule	$x^e_{1,t+1} = 0.35x^e_{1,t} + 0.65x_{t-1}$
2	WTR	Weak trend-following rule	$x^e_{2,t+1} = x_{t-1} + 0.4(x_{t-1} - x_{t-2})$
3	STR	Strong trend-following rule	$x^e_{3,t+1} = x_{t-1} + 1.3(x_{t-1} - x_{t-2})$
4	LAA	Anchoring and adjustment rule	$x^e_{4,t+1} = 0.5\bar{x}_{t-1} + 0.5x_{t-1} + (x_{t-1} - x_{t-2})$

Note: Symbol $\bar{x}_{t-1}$ denotes simple average of all observations up to period $t-1$.

normally distributed random variables with zero mean and standard deviations σ_π and σ_y, respectively. Non-negative parameters τ and κ represent the slopes of the IS curve and NKPC, respectively.

We now introduce the key element of the model in the context of this work – the expectation formation mechanism. The expectations about future inflation gap and output gap are formed in the following way. There are four types of agents in the economy. Each of these four sub-populations uses its own prediction strategy. An example of such a strategy is a simple trend-following rule. Once the expectations are formed, they are aggregated across the population based on the corresponding proportions of different types of agents. The number of agents utilizing a particular rule-of-thumb depends mainly on the past performance of the heuristic. The degree of persistence of the population proportions, the agents' willingness to switch between the heuristics and their memory capacity with respect to the long-run performance of the heuristics are determined by parameters which are fixed in time.

More formally, let x_t be the macroeconomic variable of interest (in our case, y_t or π_t). Assuming that we have four prediction rules, let $h \in \{1, 2, 3, 4\}$ refer to one of these heuristics. The four rules-of-thumb can then be described by Table 3.1.

Let $n_{h,t}$ be the fraction of agents following heuristic h at time t. Then the aggregate expected x at time $t+1$ is given by

$$\bar{x}^e_{t+1} = \sum_{h=1}^{4} n_{h,t} x^e_{h,t+1}. \tag{3.4}$$

The fraction of agents at time t is assumed to exhibit some degree of persistence and to depend on the relative performance of the corresponding heuristic. This can be expressed by

$$n_{h,t} = \eta n_{h,t-1} + (1-\eta) \frac{\exp(\gamma U_{h,t-1})}{\sum_{h=1}^{4} \exp(\gamma U_{h,t-1})}, \tag{3.5}$$

where $U_{h,t-1}$ is the absolute performance measure for heuristic h based on observations up to time $t-1$, γ is the so-called intensity-of-choice parameter and η is the parameter which determines the degree of persistence in the relative sizes of sub-populations. It can be thought of as the fraction of agents who decide not to change their strategy regardless of its recent performance. Note that when $\eta = 0$, we have the discrete-choice mechanism from Brock and Hommes (1997); when $0 < \eta \leq 1$, we are dealing with its modified version by Anufriev and Hommes (2012).

For those agents who decide to consider changing their prediction rule, the question is how sensitive are they to the superiority of heuristic h over the rest of the heuristics in terms of performance $U_{h,t-1}$. This is exactly what γ, the intensity-of-choice parameter, formalizes. It is non-negative, and we can observe that when $\gamma = 0$, the superiority of heuristic h does not raise its popularity among the agents at all: the relative performance of the heuristics is disregarded and all fractions are equal to $\frac{1}{4}$ (for those who have decided to consider switching their strategy). When the intensity of choice γ is high, even slightly better performance of heuristic h would make it much more favorable choice for the new prediction rule.

Finally, the absolute performance of heuristic h is evaluated as follows:

$$U_{h,t-1} = F(x^e_{h,t-1} - x_{t-1}) + \rho U_{h,t-2}, \tag{3.6}$$

where F is the forecast error and $0 \leq \rho \leq 1$ is the memory parameter which determines how much do agents value past performance of their heuristics. From the upper bound of ρ it is clear that the long-run performance of a rule-of-thumb can have 50% weight at best. The forecast error is calculated as

$$F(x^e_{h,t-1} - x_{t-1}) = \frac{100}{1 + |x^e_{h,t-1} - x_{t-1}|}. \tag{3.7}$$

The expectation formation mechanism described above holds both for the output gap and inflation gap. That is, y^e_{t+1} and π^e_{t+1} are formed in the same way, and the underlying behavioral parameters, namely ρ, γ and η, are assumed to be the same for the two variables.

The model's microfoundations are based on Kurz et al. (2013) and are spelled out in Appendix A of Hommes et al. (2019).

Table 3.2: Calibration of the Model

τ	σ_y	κ	σ_π	ϕ_π	ϕ_y	ρ	γ	η
1.00	0.10	0.30	0.10	1.50	0.00	0.70	0.40	0.90

3.2 Monetary Policy Implications

This model is interesting, because it departs from the strong assumption about expectation formation process – the agents' rationality – by introducing perhaps the most prominent expectation formation process in the LtFE literature, both in macroeconomic and financial experiments (Hommes et al. 2019; Assenza et al. 2019; Hommes et al. 2008; 2005). This changes the way the forward-looking NKM behaves significantly. Specifically, under calibration from Clarida et al. (2000) and Anufriev and Hommes (2012), the model implies different (from that implied by RE specification) optimal monetary policy for central banks which follow strict inflation targeting strategy: to stabilize inflation, such central banks should not react only to fluctuations in inflation gap, but also to the output gap fluctuations. The calibration is presented in Table 3.2.

Note that $\phi_y = 0$ here. This emphasizes our focus on the central banks which have inflation stability as their primary goal. As a warm-up, let us have a look at a realization of the behavioral model and contrast it with that of RE model, i.e. setting $y^e_{t+1} = \bar{y} = 0$ and $\pi^e_{t+1} = \bar{\pi} = 0$. Figure 3.1 depicts both of them.

Thanks to its expectation formation mechanism with backward-looking components, the behavioral model seems to generate more persistent time series with higher variance. As a consequence, HSM implies different optimal policy rule at least for central banks which have inflation stability as its primary goal. In particular, as shown in Hommes et al. (2019), even if a central bank is only interested in inflation stability, it can achieve better results if it also reacts to fluctuations in the output gap. Figure 3.2 depicts the relationship between the output gap reaction coefficient ϕ_y and inflation gap volatility as well as volatility of output gap and interest rate gap and contrasts it with RE model.

The volatility is measured using formula $v(x) = \sum_{t=2}^{T}(x_t - x_{t-1})^2$. This measure has at least one advantage over mean squared deviation, perhaps a more common volatility measure: the latter "does not distinguish between erratic behavior around the target with decreasing distance from the target and slow convergence if the absolute distance to the target is always equal"

Figure 3.1: RE and HSM Realizations

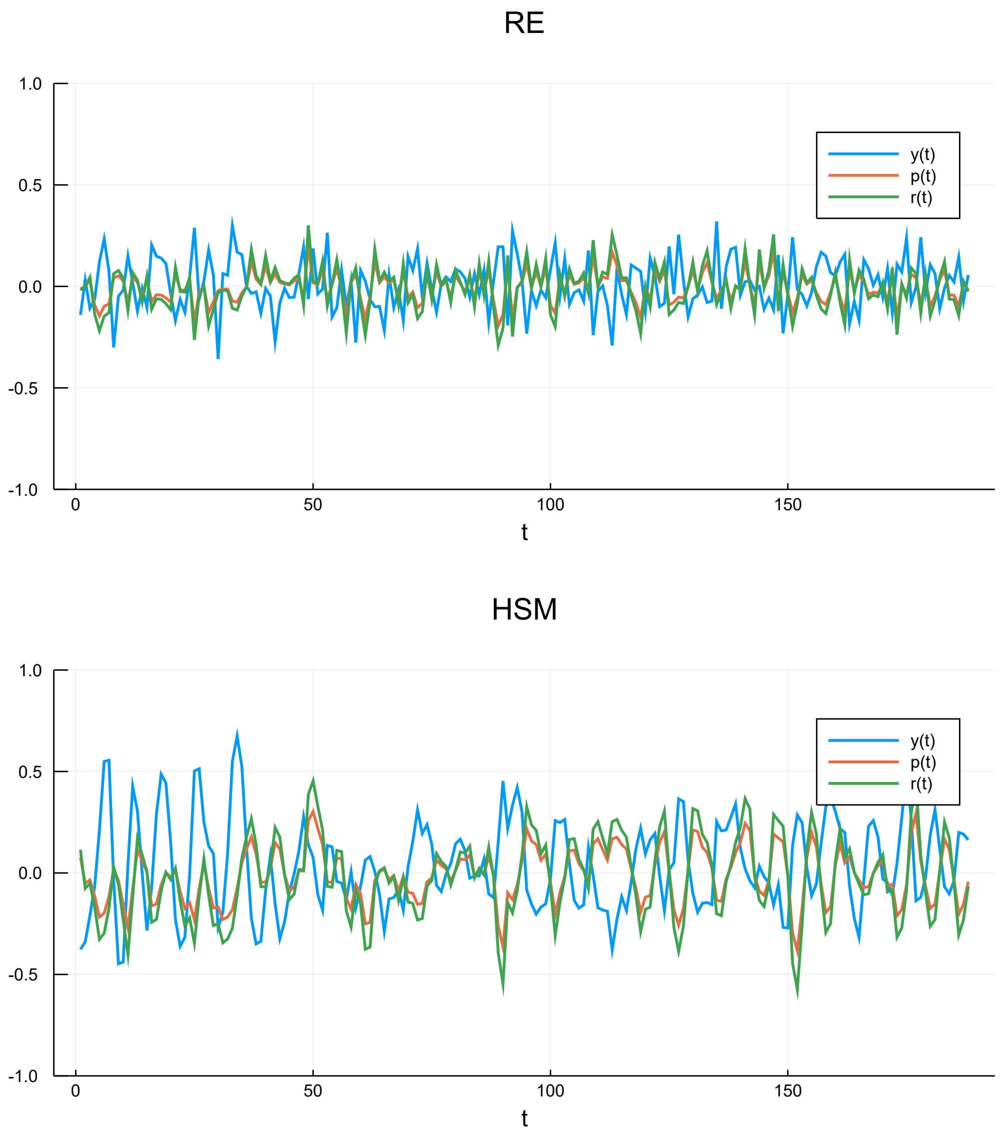
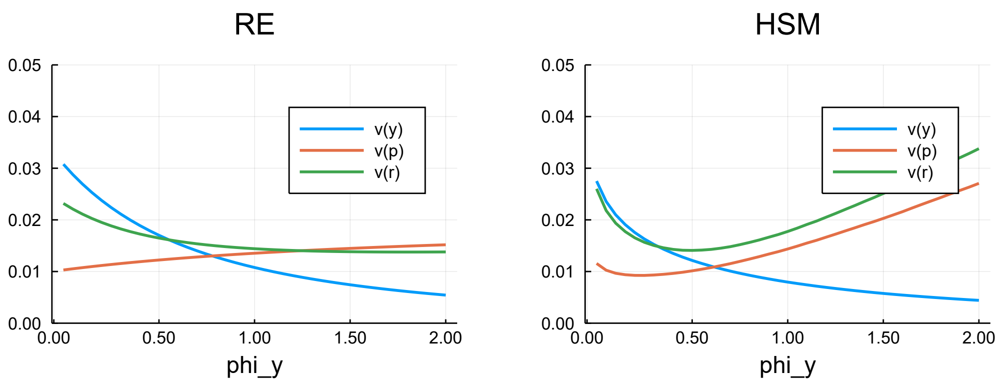


Figure 3.2: Inflation Volatility



(Hommes et al. 2019).

Turning to the discussion of different shapes of the curves from Figure 3.2, the intuition behind this difference can be illustrated by the following example from Hommes et al. (2019). Consider a situation when there is no shock to the inflation gap, but there is a positive shock to the output gap. Under rational expectations, the best thing a strict-inflation-targeting central bank can do is to idle. However, under backward-looking expectations, leaving the interest rate unchanged might lead to an endogenously-generated upward pressure on inflation in the future: high output gap today will fuel expectations about future output gap, which in turn will pull the inflation gap along. Although reacting to the initial output shock by rising the interest rate would destabilize inflation, too, this would well pay off in the near future, when the resulting modest expectations would not disrupt the steady state too much, so the total inflation volatility is lower.

Figure 3.2 is produced using $T = 188$ with first $D = 100$ time periods being discarded as the burn-in period; the number of simulations for each ϕ_y is 10,000; inflation volatility is summarized by the simple average of the 10,000 values. Other combinations of D and T produce similar results. Other volatility metrics as well as different calibrations produce qualitatively very similar results, too (see Appendices in Hommes et al. 2019).

It is worth noting that comparable results are obtained by De Grauwe (2011) who exercises his specification of behavioral expectations in the NKM framework in a similar way. His model is not based on the results of macroeconomic or financial LtFE, nor is it microfounded, but nonetheless, it is estimated several times (cf. Jang and Sacht 2016; Kukacka et al. 2018; Jang and Sacht 2019). In contrast, the model described here is fully microfounded and is grounded on the most recent developments in macroeconomic and financial LtFE literature, which makes it more interesting, perhaps more promising, but certainly not easier to estimate.

Chapter 4

Methodology

This chapter describes the way we bring the HSM to macroeconomic data. The first sub-section introduces the SMM, while the second one is devoted to Monte Carlo experiments.

4.1 The Simulated Method of Moments

We estimate the model from Hommes et al. (2019) via the method of moments. Since the moments for HSM can hardly be derived analytically, we have to approximate them by the moments calculated from a large number of simulations and try to bring them as closely as possible to its empirical counterparts. This is the main idea of the SMM.

The estimation procedure described below is relying on the exposition from Franke (2019), which is in turn based on Lee and Ingram (1991) and Duffie and Singleton (1993). Other classic papers on this topic include McFadden (1989) and Pakes and Pollard (1989).

Let $\{x_t^{emp}\}_{t=1}^{T}$ be the empirical time series, where T is the total number of periods, and x_t^{emp} is in our case a 3-tuple of 3 macroeconomic variables: y_t^{emp}, π_t^{emp}, r_t^{emp}. In defining our moments of interest, we follow Jang and Sacht (2019), Jang and Sacht (2016) and Franke et al. (2015) who solve similar problem and consider a set of 78 basic moments: variances, auto-covariances and cross-covariances. The lag length is 8, i.e. 2 years backward. The set of 78 moments thus consists of 3 variances, 3 contemporaneous cross-covariances, 24 auto-covariances (3 variables times 8 lags) and 48 lagged cross-covariances (2 times 8 for each of the 3 pairs of variables). Translating this into the language of moment functions, let m_i denote one of the 78 moment functions. We illustrate the way these moment functions work by defining moment function for cross-covariance between y_t and π_{t-4}. Assuming that this cross-covariance is, for example, 32^{nd} moment out of 78, for $t \in \{5, ..., T\}$ m_{32} is defined as

$$m_{32} : t \mapsto (y_t^{emp} - \bar{y}^{emp})(\pi_{t-4}^{emp} - \bar{\pi}^{emp}), \tag{4.1}$$

where $\bar{y}^{emp} = \frac{1}{T}\sum_{t=1}^{T} y_t^{emp}$ and $\bar{\pi}^{emp} = \frac{1}{T}\sum_{t=1}^{T} \pi_t^{emp}$ are averages over the whole sample.[2] The resulting 32^{nd} empirical moment is then given by the average of all available values of the moment function, i.e.

$$m_{T,32}^{emp} = \frac{1}{T-4}\sum_{t=5}^{T} m_{32}(t). \tag{4.2}$$

The rest of the 78 empirical moments are defined similarly. Once all of them are calculated, we end up with the vector of empirical moments $m_T^{emp} = [m_1; ...; m_{78}]$.

Now that we have calculated m_T^{emp}, we want to find such parameter vector $\theta = [\tau; \sigma_y; \kappa; \sigma_\pi; \phi_\pi; \phi_y; \rho; \gamma; \eta]$ that, under this calibration, the model generates the same or almost the same moments as the empirical time series. That is, we want to minimize the distance between simulated moments and its empirical counterparts with respect to θ. But before we formalize this aspiration into a metric, we need to introduce the way we calculate simulated moments.

Given parameter vector θ, we generate N simulations, each with length T (after initial D periods are discarded). For $n \in \{1, ..., N\}$ we denote the n^{th} simulation as $\{x_t^n\}_{t=1}^{T}$. For each of these N simulations we calculate the vector of 78 moments in the same way as in case of empirical realizations. We then summarize N vectors of moments by taking their average over these N simulations. This is our vector of simulated moments $m_T^{sim}(\theta)$ for given θ.

Recall that our goal is to minimize the distance between m_T^{emp} and $m_T^{sim}(\theta)$. Having introduced both vectors, we can define the metric which summarizes the distance between them. We consider the sum of squared differences as our objective function. More formally, we are looking for such $\theta \in \Theta$ that expression

$$J(\theta) = \sum_{i=1}^{78} w_i (m_{T,i}^{sim}(\theta) - m_{T,i}^{emp})^2 \tag{4.3}$$

is minimized; where w_i is the weight attached to the i^{th} moment. To attach

[2]Note that by supplying a moment function with time index t, we implicitly provide it with the whole realization $\{x_t^{emp}\}_{t=1}^{T}$.

greater importance to deviations of more stable moments, it is common to set w_i to reciprocal of the standard error of the i^{th} empirical moment.[3]

Unfortunately, we do not know the true standard errors of the moments, as we have only one realization with rather small number of observations. There are several ways how to estimate standard errors of the moments from one realization, but we stick to the bootstrapping procedure suggested in Franke (2019) and justified by the fact that he has rather small sample size $T = 190$, which is almost the same sample size as ours. The procedure is fairly straightforward.

Returning to our example with covariance between y_t and π_{t-4}, we sample with replacement $T - 8$ (because the lag length is 8 in our case) indices from the index set $\{9, ..., T\}$ and obtain the new set of indices $\{t_1, ..., t_{T-8}\}$. We then calculate the bootstrapped moments similarly as the empirical ones – just take the average of the values of a given moment function across the time periods:

$$m^b_{T,32} = \frac{1}{T-8} \sum_{k=1}^{T-8} m_{32}(t_k), \tag{4.4}$$

where b denotes the iteration of bootstrapping. The total number of repetitions is $B = 5000$. The standard errors of the moments are then calculated as the standard deviations of the moments over all repetitions B.

4.2 The 4-round Cross-Validation

Now we have an objective function which takes θ as argument and returns the summary of distance between average moments generated by the model under θ and its empirical counterparts. We minimize it two times: once under assumption that the true data-generating process (DGP) is HSM and once under assumption that it is RE. It is worth noting that when calculating simulated moments for RE, we make use of its analytical tractability and set all of its auto-covariances and non-contemporaneous cross-covariances to zero.

Our basic null hypothesis is that $J^{RE} \leq J^{HSM}$. In case we do not reject this hypothesis, we would conclude that behavioral model fits macroeconomic data

[3]Strictly speaking, multiplying each term in Equation 4.3 by w_i is just a special case of a more general procedure – multiplying the vectors of differences between the corresponding moments by the reciprocal of the covariance matrix of the moments. This approach is described in more detail, for example, in Jang and Sacht (2019), but because of problems associated with the sample size – as is the case of our work – they used the diagonal matrix anyway.

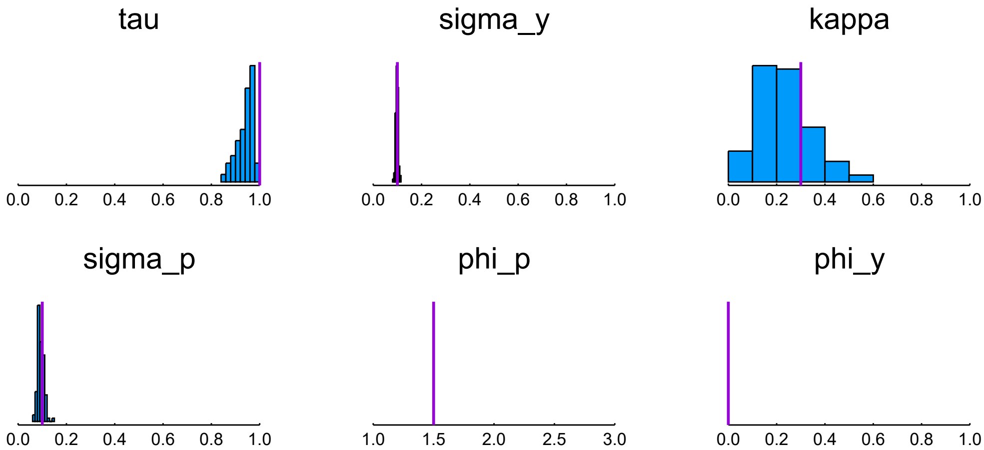
not better than RE model. Thus, monetary policy discussions from Chapter 3 are rather premature. Although outperforming the purely forward-looking RE in terms of fitness to macroeconomic data model does not seem to be a big challenge, it is the first, necessary step which has to be executed. And assuming that it outperforms RE in terms of fitness to the data, the question is how do the resulting point estimates of the parameters look like; in case the estimates are completely different from the calibration used in Chapter 3, the question is whether the policy implications based on the updated calibration still hold.

Before we bring both models to the data, we want to make sure that the procedure described above is able to discriminate between the two models. One might have a feeling that HSM is a more general model than RE – that under certain calibration of behavioral parameters (small memory and persistence parameters and large intensity-of-choice parameter, for instance) HSM fits the data not worse than RE, even if the true DGP is RE. Although the difference between the resulting J^{HSM} and J^{RE} in such case is expected to be rather small, it seems fruitful to analyze this issue more rigorously.

One way to do that is to generate a large number of realizations of RE model and 'estimate' each of these realizations twice: one time assuming that the true model is RE and one time assuming that the true DGP is HSM, and then compare the corresponding J_c^{RE} and J_c^{HSM}, where c denotes the random seed used to generate a particular realization of RE. Ideally, we want to have $J_c^{RE} < J_c^{HSM}$ for all $c \in 1, ..., C$, where C is the number of such different realizations, i.e. the number of different random seeds. In effect, we also obtain a glimpse of sample variance of our procedure for RE, when assuming that DGP is RE. That is, we assess the capability of our estimation procedure to recover the pseudo-true parameters. Ideally, in all C cases the estimates of the parameters are equal to the pseudo-true values. We can do the same for HSM: generate C realizations of HSM and 'estimate' each of them with HSM – to assess the sample variance of our estimator – and with RE – to make sure that our procedure discriminates well.

The 4-round cross-validation described above is inspired by Kukacka et al. (2018) who estimate behavioral NKM from De Grauwe (2011) via the simulated maximum likelihood method and assess the capability of their procedure to recover pseudo-true parameters of the underlying model. A similar procedure is also employed by Franke et al. (2015) as a proxy for confidence intervals for the resulting parameter estimates.

Figure 4.1: Results of Simulations: RE←RE



Note: The bold vertical lines indicate what are the true parameters of the underlying DGP. Our constrained optimization is supplied with the search range from Kukacka et al. (2018). For each parameter, the lower and upper bounds of the search range are reflected by the leftmost and the rightmost values on the x-axis, respectively.

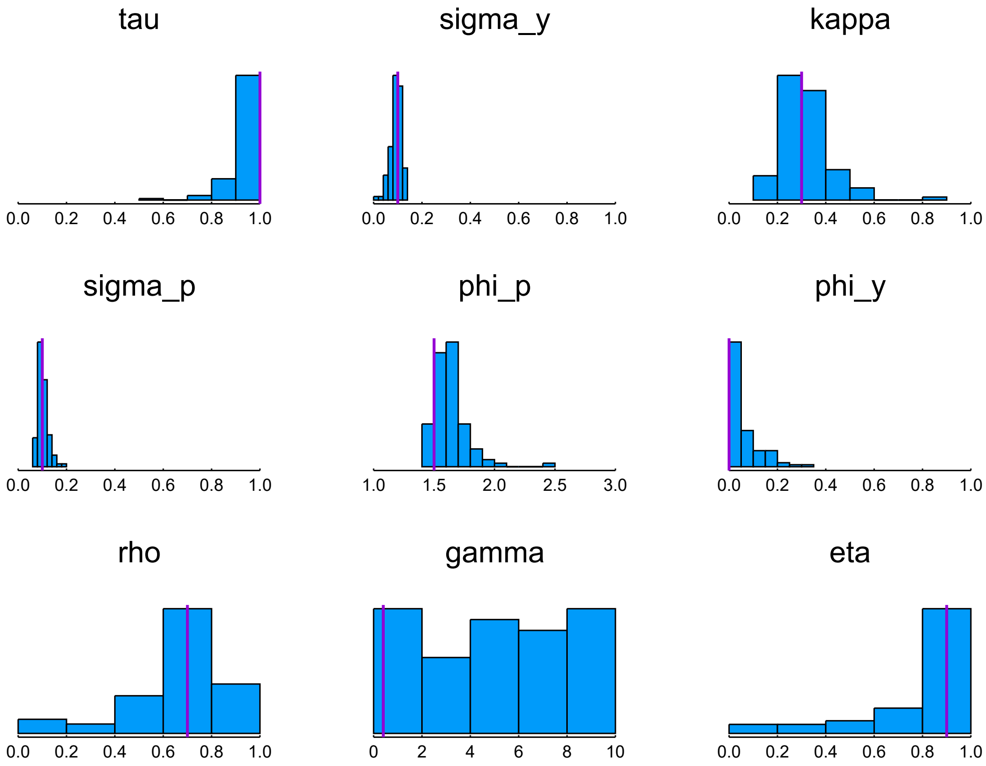
We set C equal to 100 and use the calibration from Chapter 3 with pseudo-true $\phi_y = 0.00$. Also, we follow Kukacka et al. (2018) and Hommes et al. (2019), among others, and set $\nu = 0.99$ and do not estimate it at all. As a warm-up, let us have a look at the capability of our procedure to recover the pseudo-true parameters of RE under assumption that the true DGP is RE. Figure 4.1 presents the histograms of the estimates of the pseudo-true parameters.[4]

Our estimator seems to be downward biased in case of τ. Estimation of κ does not seem to be an easy task for our procedure, either. In addition to being downward biased (even though the pseudo-true value is well inside the search range), it seems to have the highest variance among all parameters. On the other hand, parameters $\sigma_y, \sigma_\pi, \phi_\pi, \phi_y$ are recovered by our estimation procedure easily with estimators of ϕ_π and ϕ_y being remarkably sharp. This is in contrast with results of similar exercise from Kukacka et al. (2018) who report that their SMLE struggles to recover the true parameters in case of forward-looking RE model. So, we can be fairly proud of our procedure in this respect.

Let us now shatter our confidence in the estimation procedure by looking at the same graph for HSM. Figure 4.2 provides a glimpse on the variance and bias

[4]All main computations in this work are done using the Julia Language, v1.3.0 (`https://julialang.org/`). On the optimization side, we utilize the package called 'BlackBoxOptim.jl' (`https://github.com/robertfeldt/BlackBoxOptim.jl`, accessed on October 19, 2019) with 'de_rand_1_bin', one of its differential evolution algorithms, as our workhorse optimization method. More details can be found in Online Appendix D.

Figure 4.2: Results of Simulations: HSM←HSM



Note: The bold vertical lines indicate what are the true parameters of the underlying DGP. The search range for behavioral parameters is clear in case of memory ρ and population persistence η – it is interval between 0 and 1; the search range for the intensity of choice γ as well as the NKM parameters is taken from Kukacka et al. (2018). For each parameter, the lower and upper bounds of the search range are reflected by the leftmost and the rightmost values on the x-axis, respectively.

of the estimators of behavioral parameters as well as NKM parameters under heterogeneous expectations.

Under behavioral expectations our procedure seems to be able to recover parameters τ and κ slightly better than in case of RE; we may even hope that estimators of these parameters are unbiased in case of HSM. The estimators of standard deviations of the shocks seem to have higher variance in case of HSM, but are still rather unbiased. What might be disappointing for us is the distributions of estimates of reaction coefficients ϕ_y and ϕ_π. The estimators of both parameters seem to be upward biased. Regarding the behavioral parameters, perhaps the most striking observation is total impotence of the estimator of γ to recover the pseudo-true value of the parameter. This is in high contrast with Kukacka et al. (2018) whose SMLE procedure recovers the pseudo-true value of γ with remarkable precision. In addition, their estimators of ϕ_y and ϕ_π seem to work better under bounded rationality than under RE. On the other hand, they find the estimator of κ to be working much better under behavioral expectations, similarly to what we see here. Performance of estimators of the rest of the behavioral parameters leaves much to be desired: although none of them seems to be completely helpless, as in case of γ, both demonstrate rather high variance.

Anyway, the main goal of this work is to bring the HSM to macroeconomic data and evaluate its empirical performance by comparing it to that of RE model. Our null hypothesis is that the true DGP behind the realized macroeconomic time series for the euro area is RE. More formally, we reject our null hypothesis if and only if $J^{RE} > J^{HSM}$. But when testing hypotheses, it is critical to know, or at least assess, what is the p-value associated with the test and what is the probability of type II error.

Before we provide the results of such assessment, it is good to make up our minds regarding the painfulness of the two probabilities. We argue that rejecting the null hypothesis when in fact it is true would be more painful in our case. That is, rejecting the hypothesis that RE model is the true driver behind the dynamics of macroeconomic variables in the euro area and recommending the policy makers of the ECB to revise their modeling toolkit in favor of HSM when in fact RE is the true model is not the impact we want to make. Surely, reviewing one's assumptions about the way the economy works is never a bad thing, but we aim at being conservative here.

Unfortunately, as discussed above, the scenario when we falsely conclude that HSM is the true DGP is more likely than the other way around, as HSM might seem a more general model than RE. Before we face the results of perhaps the most important round of our cross-validation, let us have a look at the outcome of assessment of probability of type II error. Figure 4.3 depicts the fitness measures for the two competing models J^{RE} and J^{HSM} when the true DGP is HSM.

The fitness measure of the pseudo-true model J^{HSM} does have lower values than J^{RE} in all $C = 100$ cases. This is all that can be desired. The average relative difference between the two measures is 21.14. It is worth noting that the ability of RE to recover parameters does not seem to have deteriorated significantly: only standard deviation parameters σ_y and σ_π seem to overestimate the pseudo-true values.[5]

[5]Details, mainly in graphical form, can be found in Appendix B.

Figure 4.3: Results of Simulations: Fitness Measures for HSM←HSM and HSM←RE

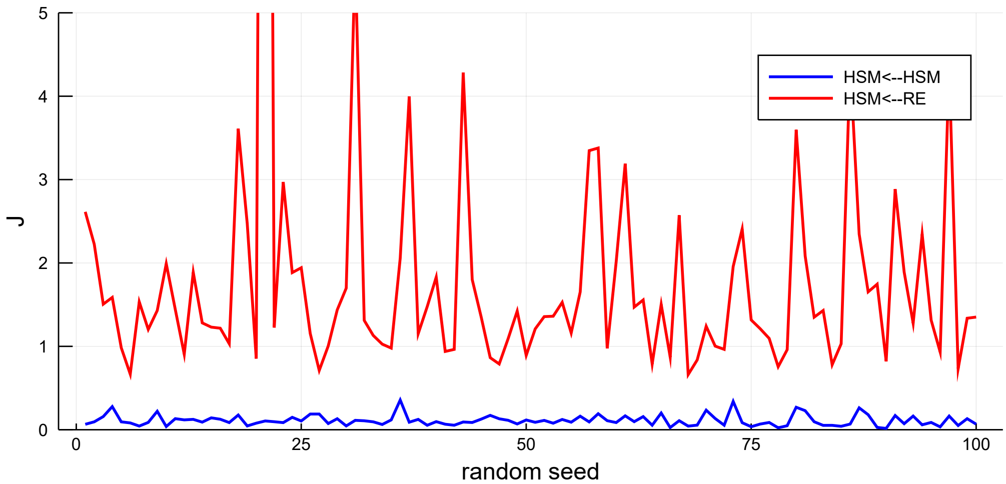
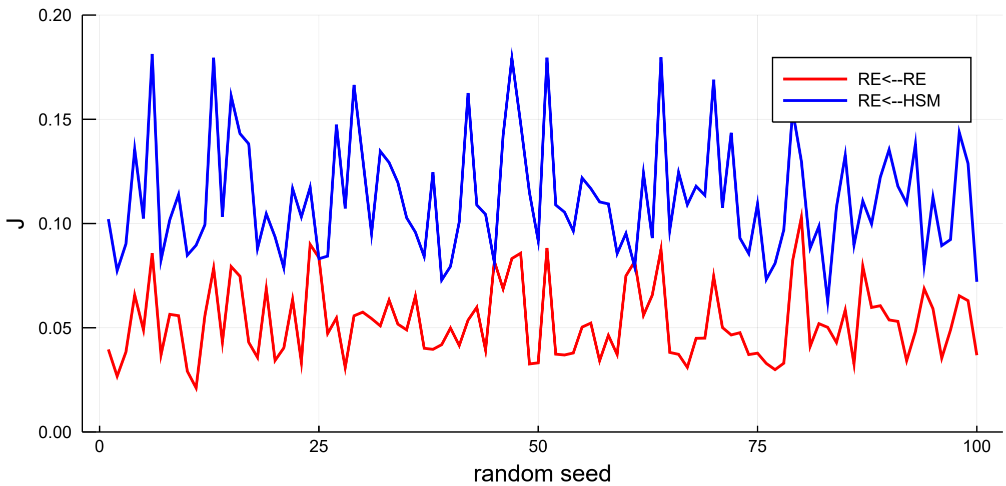


Figure 4.4: Results of Simulations: Fitness Measures for RE←RE and RE←HSM



Now that we know that type II error is unlikely to happen in our case, we can finally gauge the risk of committing a type I error. Figure 4.4 illustrates the behavior of J^{HSM} and J^{RE} across different random seeds c when the underlying true model is RE.

The difference between the corresponding fitness measures is much smaller than in case of type II error: the average relative difference $\frac{J^{HSM}}{J^{RE}}$ is only 2.28. The two curves even seem to be tangent three times. In fact, in all of these cases J^{HSM} is less than J^{RE}, although only marginally. Our estimate of p-value is thus $\frac{3}{100} = 3\%$. In addition to that, we can conclude that if empirical data produce almost identical J^{HSM} and J^{RE} then it is probably the case that the true model is RE, because should HSM be the true model, J^{HSM} would be much lower than J^{RE}, as suggested by Figure 4.3. Although we would formally reject our null hypothesis in the former case, we would suspect that we have committed a type I error. Another feature that would indicate that we have falsely rejected RE is having the estimate of γ equal or almost equal to its upper bound, 10.[6]

In summary, given our main goal – comparing the empirical fitness of HSM to that of RE – the estimation procedure described in this chapter seems to be appropriate enough: the estimated probability of type II error is 0%, and the estimated probability of type I error is 3% with reservation that we would probably know that we are about to reject RE falsely, and so the probability of type I error is practically almost 0%. Thus, the bottom row of the table with estimation results is going to be interesting and fairly reliable reflection of empirical fitness of the two competing models. Regarding the parameter inference, we are not expecting our procedure to recover all true parameters extremely accurately, but we do have some glimpse on bias and variance of our estimators, so the upper rows of the table with results are not going to be completely uninformative, too.

[6]Histogram for this as well as the rest of parameters can be found in Appendix B.

Chapter 5

Results

This chapter is devoted to empirical application of the methodology described in Chapter 4. First, the macroeconomic data are introduced. Then, the results of the estimation are presented. Finally, optimal monetary policy recommendations discussed in Chapter 3 are stress-tested.

5.1 Empirical Application

We follow Kukacka et al. (2018) and retrieve the euro area macroeconomic time series from the Euro Area Business Cycle Network. In particular, we utilize the 18$^{\text{th}}$ update of the area wide model database which covers the time window from 1970Q1 to 2017Q4 on the quarterly basis (for a detailed description of the dataset see Fagan et al. 2001).[7] For our purposes, we need three variables: YER (real GDP), YED (GDP deflator) and STN (the short-term nominal interest rate). Once GDP and its deflator are transformed to year-over-year difference, we filter out the trend in all three variables using the Hodrick-Prescott filter with usual smoothing parameter $\lambda = 1600$. The resulting time series are plotted in Figure 5.1.[8]

Now we are all set to bring the HSM to macroeconomic data. The results of the estimation are presented in Table 5.1.[9] Optimization constraints are reflected in the angle brackets.

The ratio $\frac{J^{RE}}{J^{HSM}}$ is equal to 22.45, which is slightly greater than 21.14, the

[7]Available at: `https://eabcn.org/page/area-wide-model` [Accessed on December 2, 2019].

[8]See Appendix C for the same graph before the application of the Hodrick-Prescott filter.

[9]Note that the initial search range for parameter σ_y was set to <0, 1>. But because RE produced estimate of σ_y equal to 1.00, the upper optimization constraint was raised to 2. The results of the original estimation are presented in Appendix C.

Figure 5.1: Empirical Time Series

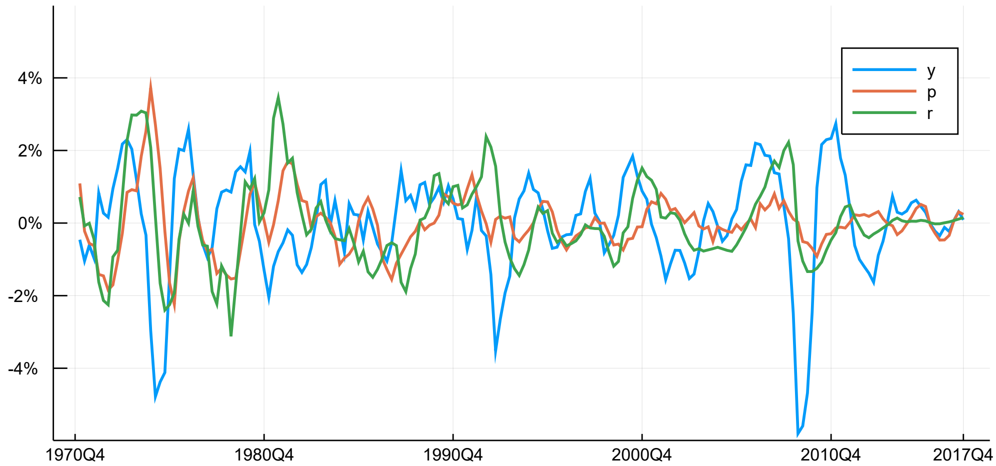


Table 5.1: Results of Estimation

Parameter	HSM	RE
τ	0.33	0.95
<0, 1>		
σ_y	0.21	1.86
<0, 2>		
κ	0.05	0.21
<0, 1>		
σ_π	0.00	0.83
<0, 1>		
ϕ_π	2.01	1.62
<1, 3>		
ϕ_y	0.43	0.28
<0, 1>		
ρ	0.00	-
<0, 1>		
γ	7.04	-
<0, 10>		
η	0.53	-
<0, 1>		
J	5.96	133.83

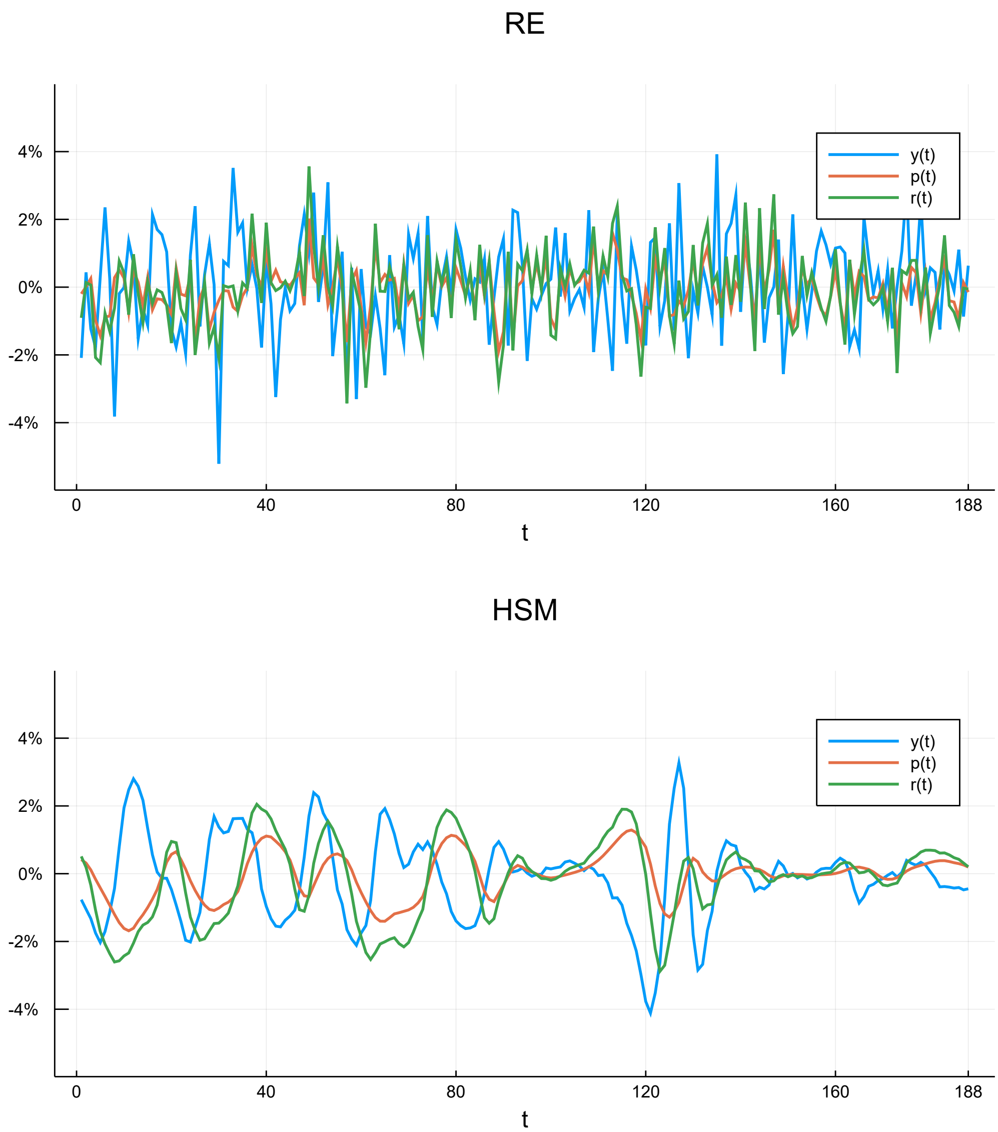
corresponding Monte Carlo benchmark from Chapter 4. This indicates that the behavioral model does fit the macroeconomic data better than the RE model. More formally, the null hypothesis from Chapter 4 is rejected on the 3% level. In other words, the true DGP behind the macroeconomic time series for the euro area is rather HSM than RE. This result is not surprising: fitness measures from Jang and Sacht (2019) and Kukacka et al. (2018) tend to favor behavioral models, too.

Regarding the coefficients, it might be tempting to proclaim that our study confirms the results from Kukacka et al. (2018) whose SMLE procedure manages to pin down the intensity-of-choice parameter γ to 7.01, but Figure 4.2, namely, the flat shape of the distribution of estimates of γ, suggests that one should be cautious in this respect. The estimate of another behavioral parameter, memory ρ, is almost zero, which supports its insignificance reported by Jang and Sacht (2016) and implied by Kukacka et al. (2018), among others. This is at odds with the calibration by Hommes et al. (2019) who set it to 0.7. Finally, the population persistence parameter η is estimated to be 0.53, which is far from what is suggested by Hommes et al. (2019), too, but can hardly be compared to empirical studies by Jang and Sacht (2016), Jang and Sacht (2019) or Kukacka et al. (2018), either, as they implicitly, by design of the discrete choice mechanism, assume that it is equal to zero. Notably, the estimates of all behavioral parameters depart from the calibration by Hommes et al. (2019) in the same direction as in Bao et al. (2020) who test the robustness of the results of LtFE by Hommes et al. (2008) with respect to a larger number of participants.

The estimates of the structural parameters τ, κ, ϕ_π and ϕ_y as well as standard deviations of the error terms σ_y and σ_π from this work are generally closer to those from empirical studies by Jang and Sacht (2019) and Kukacka et al. (2018) than to the calibration by Hommes et al. (2019), keeping in mind that the estimators of standard deviations of shocks are very likely downward biased in Kukacka et al. (2018). This is especially true in case of the central bank's reaction coefficients ϕ_π and ϕ_y, whose true values are likely to be less than what is reported in Table 5.1, because its estimators seem to be upward biased.

Interestingly enough, the estimates of the parameters of RE model are quite close to the calibration by Hommes et al. (2019), except for the standard deviations which are much greater here. In contrast, the behavioral model does not need big exogenous shocks to explain the covariance profiles of the empirical data: its endogenous expectation formation process seems to be sufficient to

Figure 5.2: Results of Estimation: RE vs. HSM

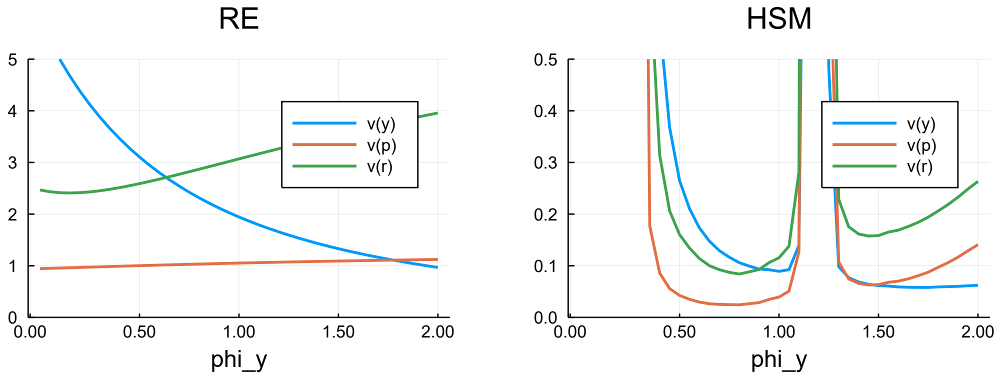


account for given features of macroeconomic time series. Let us have a look at the examples of realizations of these two models under the calibration from Table 5.1. Both are depicted in Figure 5.2.[10]

All in all, the behavioral model outperforms the RE model in terms of fitness to macroeconomic data, both formally and visually. However, the set of parameters which 'secured victory' over RE is very far from being equal to that from Hommes et al. (2019). And although the estimates look plausible, before declaring that HSM has passed the empirical validation and that the ECB

[10]The underlying population fractions for HSM are contained in Appendix C.

Figure 5.3: Inflation Volatility: A Robustness Check



should revise its optimal monetary policy rule, it should be verified that the optimal policy implications discussed in Chapter 3 hold.

5.2 Monetary Policy Exercise

Ideally, HSM with calibration from Table 5.1 should be able to reproduce Figure 3.2, at least qualitatively. That is, inflation volatility should still be decreasing for low ϕ_y and increasing for ϕ_y greater than, say, 0.5. Figure 5.3 depicts the results of the same exercise as in Chapter 3.

The setup (volatility measure, number of simulations, etc.) is the same as that used for producing Figure 3.2. The only difference is the underlying calibration of the models. The curves for HSM look a bit weird. Before we discuss its ups and downs, let us consider the curves for RE.

This version of RE model generates very similar curves in terms of dynamics. That is, $v(y)$ decreases as ϕ_y increases; $v(p)$ demonstrates the opposite; and $v(r)$ is decreasing for small ϕ_y. What is different is that $v(r)$ soon starts to rise, as opposed to its counterpart from Figure 3.2 where the curve is decreasing over the whole sensible range of ϕ_y. This has to do with the fact that σ_y is now much larger than σ_π, as opposed to the original calibration where both are equal.

Turning to the curves for HSM, let us first investigate the model's explosions when ϕ_y is small. Figure 5.4 shows an example of such realizations.

Not surprisingly, the most frequently used heuristic during this realization is the strong trend-following rule.[11] So, the logic here is similar as in the case of Figure 3.2: not reacting to the initial shock to the output gap fuels the

[11] See Appendix C for corresponding graphs.

Figure 5.4: An Example of HSM Explosions ($\phi_y = 0.00$)

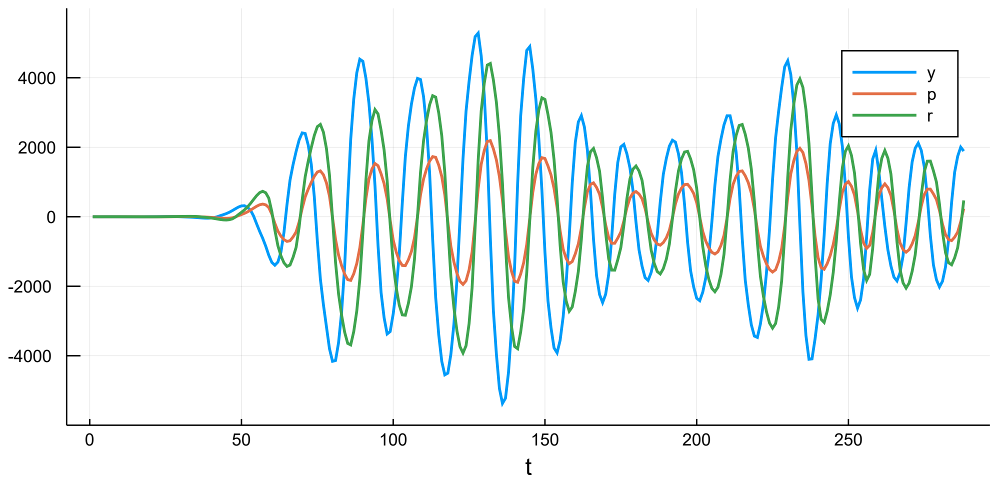
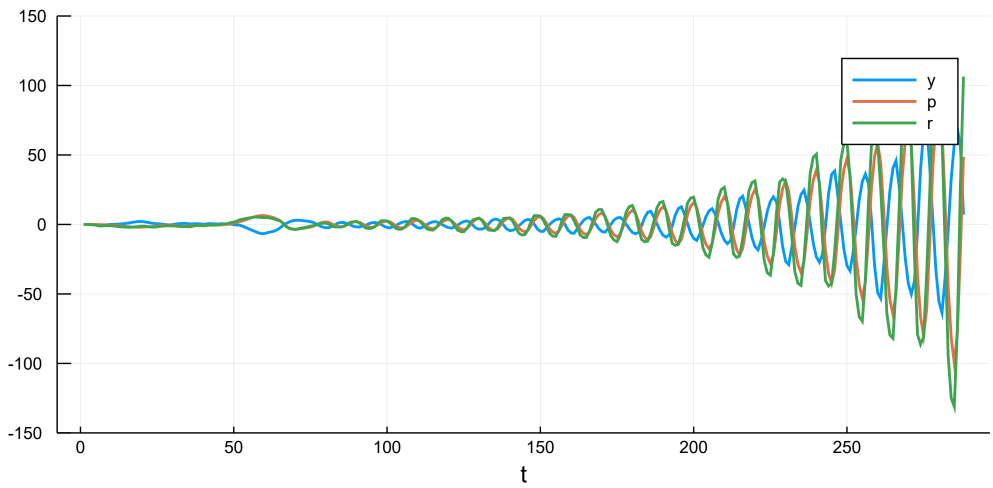


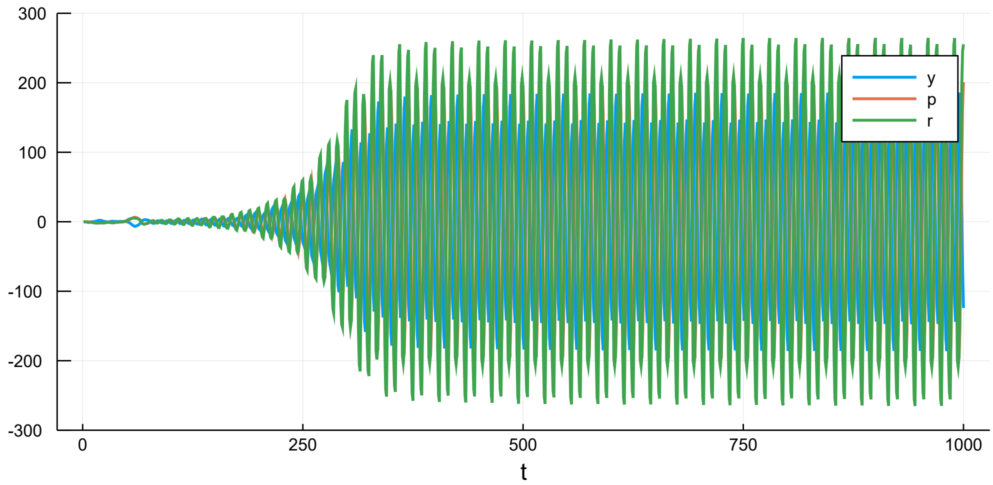
Figure 5.5: An Example of HSM Oscillations ($\phi_y = 1.20$) – Short Run



expectations about the future output gap and thus leads to a higher output gap in the future. In addition, because the calibration from Table 5.1 yields much smaller estimate of κ, only 0.05, inflation and output gap do not go hand-by-hand, and without the central bank's attention ($\phi_y = 0$) the output gap can easily 'go wild' and consequently blow the whole economy up, especially given that the central bank is having hard times bringing inflation back to its steady state, since the monetary policy channel is clogged ($\kappa = 0.05$ and $\tau = 0.33$).

Regarding the 'time warp' to the right, this one is really puzzling. Only a small fraction of all time series fall into the 'singularity' at $\phi_y \approx 1.2$. Figure 5.5 depicts one of these 'ill-starred' realizations.

Figure 5.6: An Example of HSM Oscillations ($\phi_y = 1.20$) – Long Run



This is not the way a healthy economy behaves. Although $\phi_y = 1.2$ is not the most sensible calibration, neither it is completely off. One might wonder what happens to the model in the subsequent periods. Figure 5.6 answers this question.[12]

And this is not the kind of equilibrium macroeconomic DSGE models aim at producing. This graph also emphasizes the importance of having sufficient burn-in period length B. Turning to the drivers behind these bizarre dynamics, Figure 5.7 indicates what kind of force is holding everything together.[13]

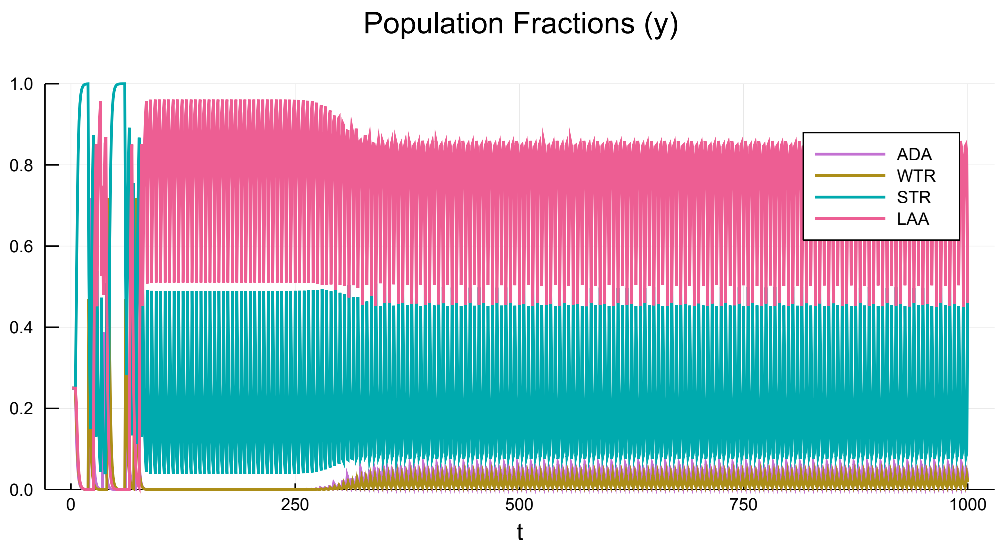
Expectations play a very important role in macroeconomics. And population fractions play an important role in expectation formation mechanism. Especially when the underlying persistence parameter η is close to one half, which is our case. Setting η to 0.7 neutralizes the 'black hole' at $\phi_y \approx 1.2$. Although it is rather a slice of a 'black ravine', as illustrated in Figure 5.8 with natural logarithm of $v(p)$ on the 'z-axis'.

Regardless, even at the epicenter of the 'singularity' at $\phi_y \approx 1.2$ most of the realizations are well-behaved. What is much more important is that the U-shape from Figure 3.2 is generally present. In a way, it is now more pronounced, since setting ϕ_y to small values would blow the economy up. Moreover, as suggested in Figure 5.8, setting any reaction coefficient close to zero might upset the whole economy, which might be relevant for all central banks.

[12]See Appendix C for a deeper look at the dynamics of the macroeconomic variables of this realization.

[13]See Appendix C for a deeper look at the dynamics of the population fractions of this realization.

Figure 5.7: An Example of HSM Oscillation ($\phi_y = 1.20$) – Population Fractions



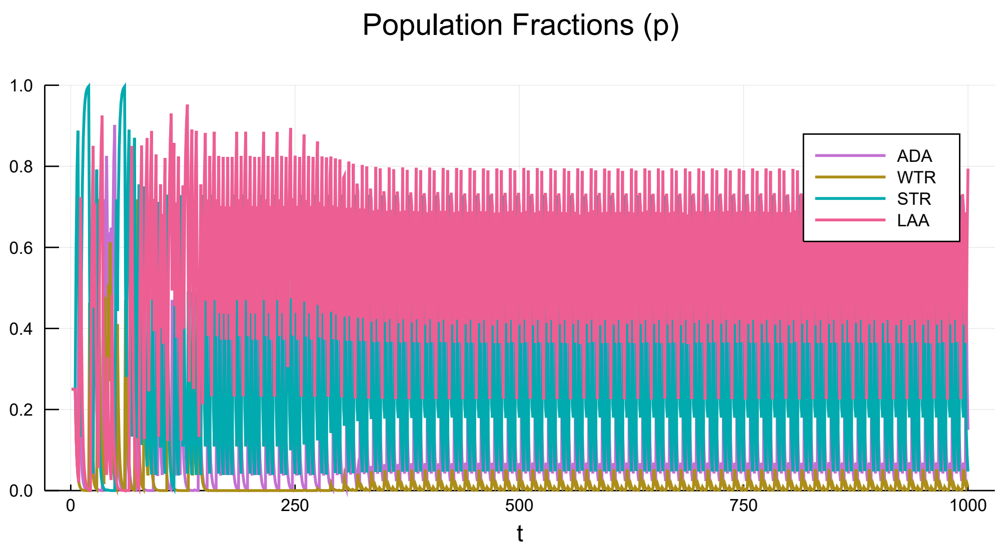
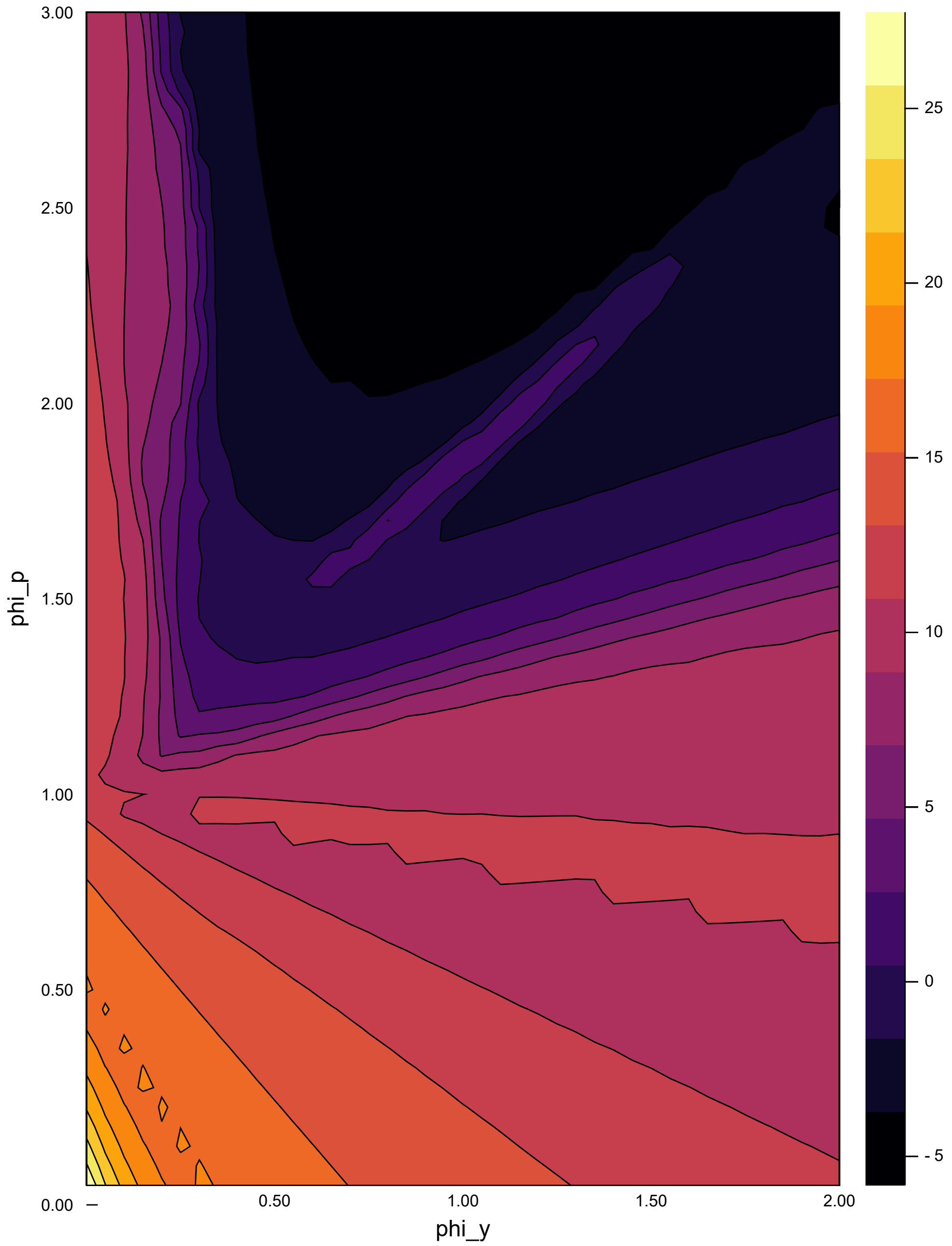


Figure 5.8: Inflation Volatility for HSM: The 'Black Ravine'



To sum up, HSM fits the macroeconomic data much better than RE. The resulting parameter estimates of the behavioral model are completely different from the calibration by Hommes et al. (2019). The inflation volatility exercise based on the new calibration shows that the optimal monetary policy implications for central banks which target inflation generally hold, with the provision that the underlying calibration might not represent an accurate reflection of the true DGP, since the precision of the underlying estimators leaves something to be desired, as is indicated in Chapter 4.

Chapter 6

Discussion

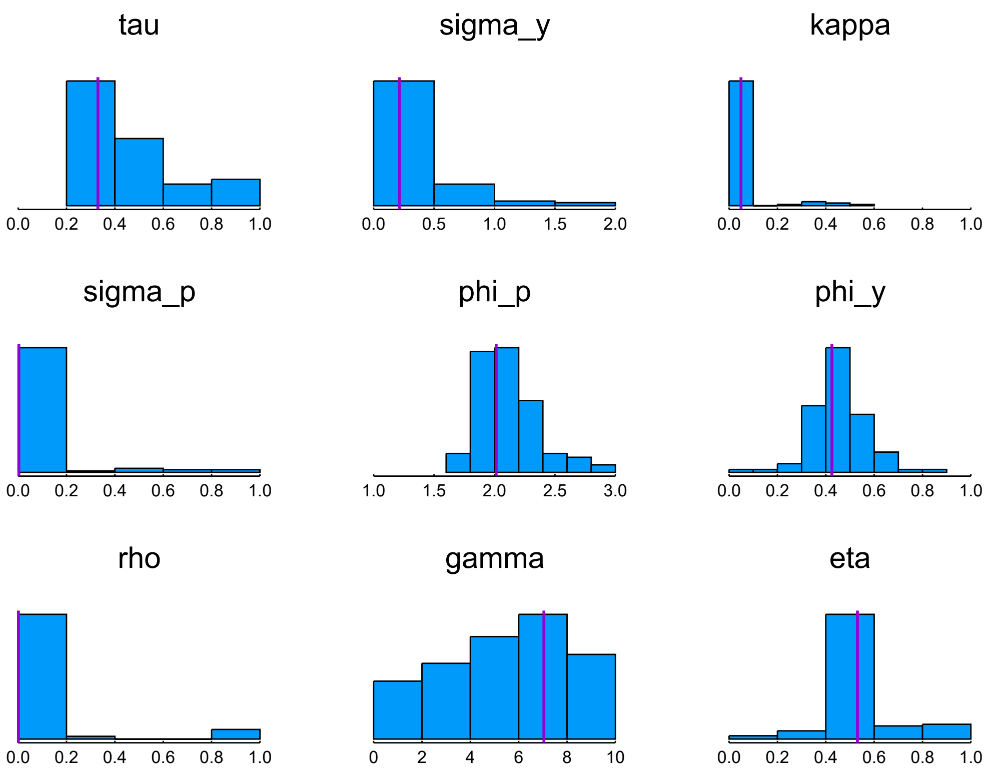
This chapter deepens the grasp of the empirical results and provides a discussion on the related issues. First, a second layer of parameter inference is added. Next, the distributions of the underlying fitness measures J are presented and the empirical results from Chapter 5 are checked against the moment-matching graph. Finally, alternative fitness measures are proposed.

6.1 A Deeper Look at the Estimation Results

When looking at the results of the 4-round cross-validation from Chapter 4 and Appendix B and the results of the estimation from the previous chapter, one might question the applicability of the former in assessment of the variance of the estimators. That is, since the true parameters of the underlying DGP seem to be in a completely different interval of the parameter space Θ than that implied by the calibration used in Chapter 4, the estimators might behave differently. We may well discover that our estimators, for example, have high variance, and this would undermine the results of the inflation volatility exercise from Chapter 5. Let us thus have a look at Figure 6.1 – the empirical counterpart of Figure 4.2 based on the estimation results from Chapter 5 and sample length $T = 188$.

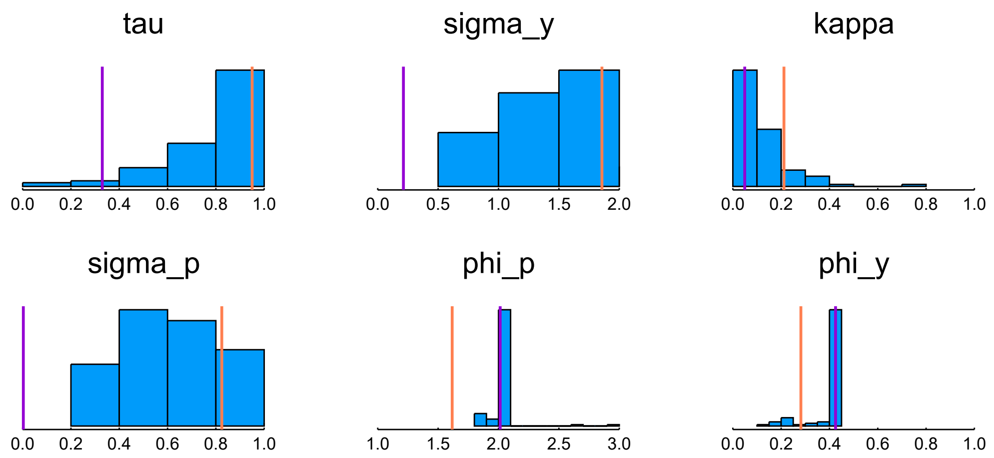
Apart from the underlying calibration, the extended search range for σ_y and the sample length T, the whole setup is the same as that used for producing Figure 4.2. If anything, the accuracy of the parameter estimators seem to have slightly improved compared to what we have seen in Chapter 4: although their variances are still non-negligible and a couple of them are somewhat biased, none of the estimators is glaringly biased, as is the case of ϕ_π and ϕ_y

Figure 6.1: Results of Simulations Based on Empirical Results: HSM←HSM



Note: Vertical violet lines locate the underlying calibration.

Figure 6.2: Results of Simulations Based on Empirical Results: HSM←RE



Note: Vertical violet lines locate the underlying calibration, while the coral lines mark the locations of the actual empirical estimates.

in Figure 4.2. Even the distribution of the estimates of the intensity-of-choice parameter γ is no longer flat.

Let us complete the second layer of parameter inference by repeating the same exercise under assumption that the true model is RE. Our hope is that the estimates are distributed around its empirical counterparts from the third column of Table 5.1 in Chapter 5. Figure 6.2 depicts the resulting distributions.

The idea of the distributions in Figure 6.2 is to show how does the estimation procedure behave if the true DGP is HSM with calibration based on empirical results and the assumption of the estimation is that the true model is RE. At first glance, the locations of the empirical estimates look very good: when the true model is HSM, the estimation procedure under assumption that the true DGP is RE does seem to produce estimates similar to what we have obtained in Chapter 5. This is especially true in case of τ. Together with the locations of σ_y, σ_π and κ, it does not seem unlikely that the true model is HSM. However, the estimates of the reaction coefficients ϕ_π and ϕ_y are completely off here: the extreme sharpness of the estimators of ϕ_π and ϕ_y under assumption that the true model is RE seems to be highly robust to the underlying DGP. That is, these parameters are consistently recovered with remarkable precision which can be seen already from Figure B.1. So, chances are the calibration from the second column of Table 5.1 – estimation results for HSM – is far from being an accurate reflection of the true parameters standing behind the empirical data, at least in case of the central bank's reaction coefficients.

This calls for fixing ϕ_π and ϕ_y in accordance with the rightmost column of Table 5.1, should the model be re-estimated. Another sound candidate for fixing is the memory parameter ρ which is consistently estimated to be zero (see Jang and Sacht 2016, among others). Finally, one can make use of the results of Kukacka et al. (2018) and fix the intensity-of-choice parameter γ to their estimate, 7.01. Anyway, the main goal of this work is to compare the empirical fits of the two models, so our primary concern is not the parameter estimates, but the fitness measures J's.

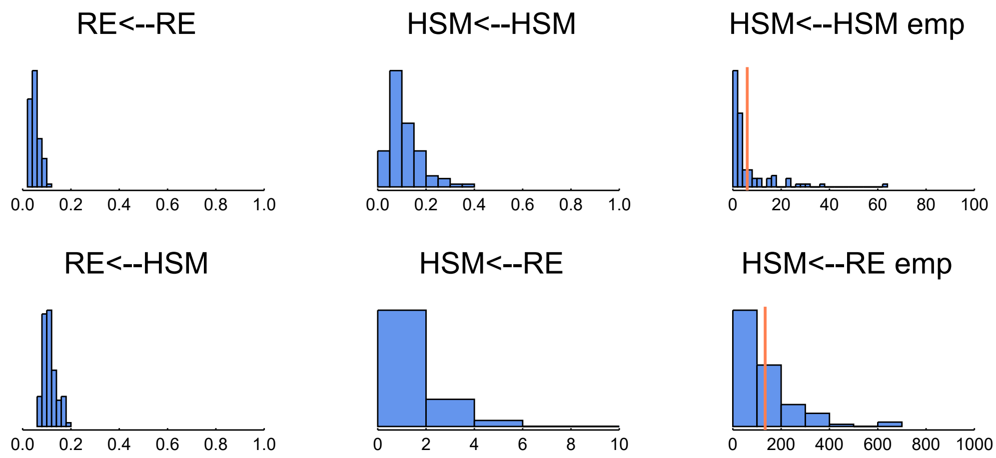
Speaking of J's, a meticulous reader might have noticed that something is uneasy with these measures already in Chapter 4 and Appendix B. In particular, what might make a clumsy impression is that only absolute and relative differences between the two metrics are considered (see Figure B.3 and Figure B.4), and its values are not even presented. Which is enough to compare the two models' fits once they are brought to the macroeconomic data, but the question is whether HSM fits the data well and RE fits the data poorly or both models are inaccurate approximations of the underlying DGP, but the RE model is even worse than the behavioral one.

Recall that the values of J are 5.96 and 133.83 for HSM and RE, respectively. In contrast, the median J for Figure 4.2 (HSM$\leftarrow$HSM) is 0.09 and 1.35 for Figure B.1 (HSM$\leftarrow$RE), which might suggest that both models fit the data very poorly. Before we investigate this suggestion, let us have a look at the distributions of the fitness measure under different setups. Figure 6.3 discloses all the six distributions of J – one for each of the sets of histograms presented in this work, including those from Appendix B.

Looking at the top middle histogram, one might conclude that $J = 5.96$ implies a remarkably bad fitness, although still much better than that implied by 133.83, the value for RE. However, when the upper right distribution enters the reasoning, it does not seem very unlikely that the true DGP is HSM with parameters from Table 5.1. Theoretically, the weighting matrix should take care of such movements in the parameter space Θ, but in practice, when some subspaces of Θ produce on average less stable time series than the other, there is no guarantee that all the candidate models are brought to the same metrics. This seems to be exactly the case of our model, where the interval around θ^* from Table 5.1 appears to occasionally generate wild realizations, as is suggested, for example, by Figure 5.8.

Luckily, we can assess both relative and absolute fits of RE and HSM to

Figure 6.3: The Underlying Fitness Measures J



Note: The histograms to the right are the underlying distributions of J based on the empirical calibration (i.e. Figure 6.1 and Figure 6.2). The vertical coral lines locate the 'realized' empirical J's.

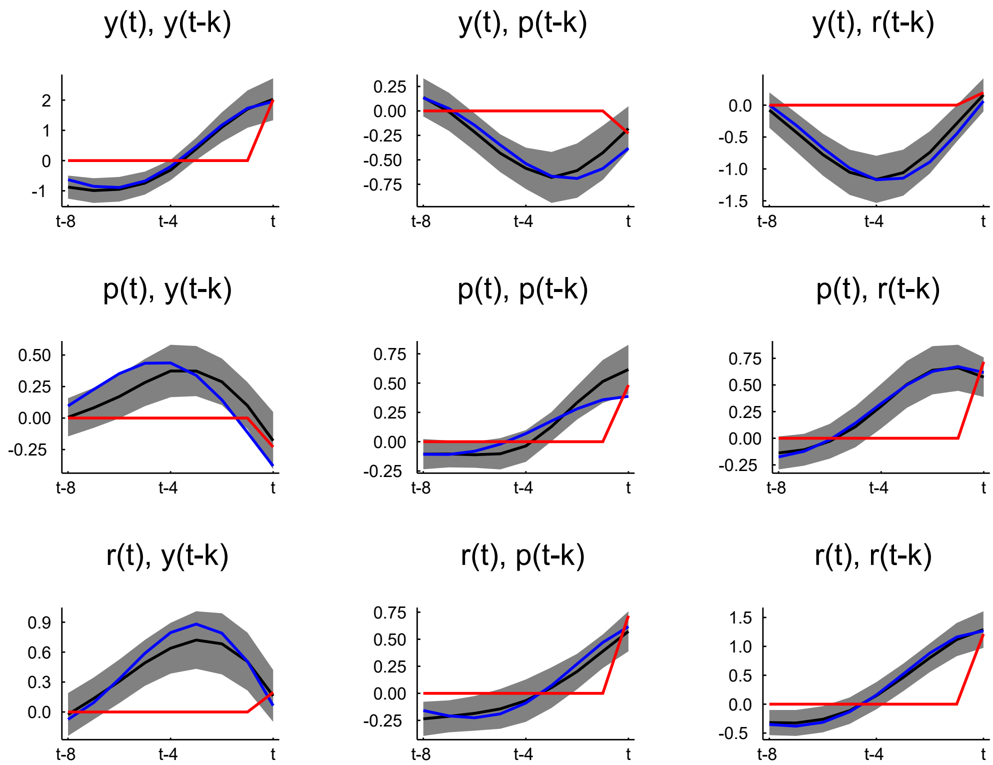
macroeconomic data in a more transparent way. Instead of solely relying on the power of the 4-round cross-validation from Chapter 4, we can simply plot the auto- and cross-covariance profiles of the two models resulting from Table 5.1 against its empirical counterparts. Figure 6.4 provides a visual check of the results obtained in Chapter 5.

Clearly, not only HSM comfortably outperforms RE in terms of fitness to macroeconomic data, but it also exhibits a remarkably good fit to the covariance profiles of the empirical dataset in general.

6.2 Alternative Fitness Metrics

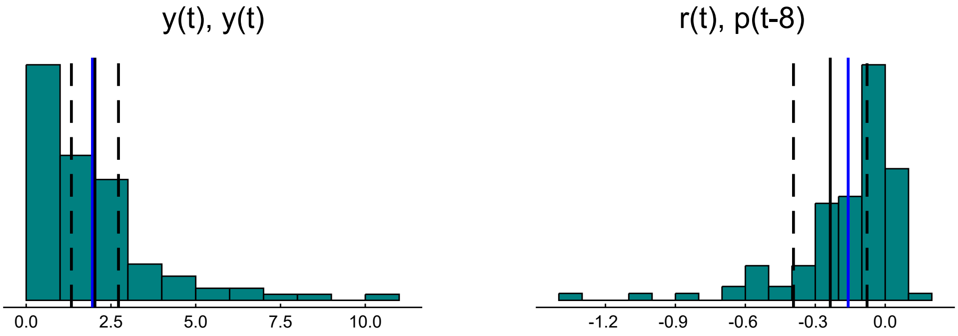
While we are on the subject of moment matching, it is worth emphasizing that the simulated moments in Figure 6.4 represent the average moments across all simulations $1, ...N$. A very meticulous reader might have already been questioning the validity of computing moments for each of these N simulations, summarizing these moments by taking their simple average and then using these average moments to calculate the distance measure J. Indeed, a legitimate concern is that it is not obvious that averaging a moment across the simulations does not distort the true picture and favor parameter vectors θ's which should otherwise be discarded during the optimization. The problem is that the distribution of the moment across the simulations does not necessarily have to be unimodal and symmetrical, as is implicitly assumed in this step.

Figure 6.4: Moment Matching: HSM vs. RE



Note: The red and the blue lines depict the simulated moments of RE and HSM, respectively. The black lines stand for their empirical counterparts; the dark ribbons indicate its confidence intervals as calculated as two times standard error of the corresponding moment.

Figure 6.5: Moment Matching: Examples of the Underlying Distributions for HSM



Note: The blue vertical lines are the corresponding averages, the black ones are its empirical counterparts with confidence bands.

And this concern is especially relevant in case of HSM, where the underlying DGP is highly non-linear.

Perhaps the best basic check we can perform here is to consider the distributions of the moments which allow HSM to outperform RE in Table 5.1. And the best outcome we can hope for is that all moments have symmetrical unimodal distributions, so the loss of information caused by averaging them across the simulations is rather negligible.

Unfortunately, this is not the case. The resulting distributions are far from looking delightfully unimodal. Although none of them look bimodal, almost all of them exhibit other unpleasant features, such as having outliers. Figure 6.5 depicts typical examples of these troublesome distributions.

Apart from being skewed or having relatively extreme values which pull the resulting average away from what might seem a more appropriate summary of the distribution, it is not uncommon to observe a moment being 'insignificant', i.e. when the moment's histogram breaches the zero threshold considerably, even though the empirical moment is not always likely to come from a distribution with zero mean, based on the confidence bands and the method of summarizing the distribution. The histogram to the right is exactly the case.

Thus, the distances between the simulated moments and its empirical counterparts are a bit warped, and therefore, fitness measure J is probably not an extremely trustworthy metric at least in some neighborhood of θ from the HSM column of Table 5.1. As a consequence, the credibility of the results of empirical application as well as the inflation volatility exercise from Chapter 5 is undermined somewhat.

There is not much one can do about it at this stage. Perhaps the simplest way to resolve this issue would be to replace the average by median. That is, when we have N atomic vectors of simulated moments, we can aggregate them into one vector by taking the median of each moment, not the average. This metric – let us call it 'J-outlier' – would take care of the outliers and skewness and supply us with a more intuitive summary of the moments. But should the distribution of the moments be, for example, bimodal, we would question the competency of our fitness measure again, although this would rather indicate more serious problems with parameter identification or optimization algorithm.

Nevertheless, to better account for potential multimodality, we may decide not to summarize the moments across the simulations at all. Instead, we can first calculate J_n for each simulation, and then aggregate them, for example, by taking their average across all $n = 1, ..., N$ simulations. This metric is inspired by Franke (2019) who employs a similar approach to calculate the p-value for his estimation. The idea is not to waive the valuable information contained in the distributions of the moments too early, but to leverage it to obtain the distribution of the sums of distances between the corresponding moments across different random seeds. This distribution can then be summarized, for example, by computing its average or median – we label them as 'J-bar' and 'J-median', respectively. The difference between these two is of a similar nature as in case of the standard metric and J-outlier: J-bar is more straightforward and computationally cheaper, but is prone to outliers. It should be noted that this approach cannot be complemented by a direct graphical check of fitness, such as Figure 6.4, which can be seen as its drawback.

In summary, the second round of parameter inference as well as a deeper look at the empirical fit and the underlying fitness measures presented in this chapter do not discredit what is reported in Chapter 5 entirely. Although there seems to be some room for a methodological 'horse race' in the future, given the results of the 4-round cross-validation and the second layer of parameter inference, none of the SMM aspects discussed here look critical in the context of this work, and so the conclusions made in the previous chapter generally hold.

Chapter 7

Conclusion

Being concerned with the gap between the theoretical soundness and empirical relevance of one of the most prominent behavioral macroeconomic models in the literature, we bring the model from Hommes et al. (2019) to empirical data using the simulated method of moments (SMM). We employ the standard rational expectations (RE) specification of the New-Keynesian model (NKM) as a benchmark and utilize Monte Carlo experiments to assess the discriminatory power of the estimation procedure. We also provide a robustness check to the implications of behavioral expectations for optimal monetary policy presented in Hommes et al. (2019) and De Grauwe and Ji (2020), among others. Finally, we discuss possible methodological caveats and propose three alternative fitness metrics.

The Heuristic Switching Model (HSM), one of the leading behavioral expectation formation mechanisms in the learning-to-forecast experiments (LtFE) literature, delivers a remarkably good fit to the auto- and cross-covariance profiles of the euro area macroeconomic time series and outperforms RE at the 3% significance level. The inflation volatility exercise based on the resulting calibration corroborates the claim that central banks which implement strict inflation targeting are better off reacting to the output gap fluctuations, on top of inflation. On the methodological side, the newly proposed metrics – a very preliminary response to one minor concern related to the fitness measure J – appear to have a potential to enhance the SMM approach in this respect and allow it to reinforce its status of a sound framework for estimating macroeconomic models with behavioral expectation formation mechanisms.

All in all, based on our empirical examination, it may be said that HSM is validated with minor conditions: the model matches empirical moments supremely well with a remark that there seems to be a scope for improvement of the underlying estimation procedure, at least in one aspect. Once it is studied more extensively and the SMM is tweaked accordingly, the model can be challenged by other microfounded behavioral specifications of NKM, such as 'fundamentalists-chartists' models by De Grauwe and Ji (2020) or Hommes and Lustenhouwer (2019). Effectively, this comparison might also shed more light on the applicability of the results of laboratory experiments to the real macroeconomic environment, as the more parsimonious 'fundamentalists-chartists' setup has a weaker support in the LtFE literature than the more sophisticated HSM, but may well outplay the latter when applied to the real macroeconomic data. Either way, this work adds a considerable piece of evidence for behavioral heterogeneity in macroeconomic expectations in the euro area, which seems relevant for the ECB already now, although the final specification of the potential behavioral replacement of RE in its modeling toolkit is to be determined by further research.

Bibliography

Anufriev, M. and Hommes, C. (2012), 'Evolutionary Selection of Individual Expectations and Aggregate Outcomes in Asset Pricing Experiments', *American Economic Journal: Microeconomics* **4**(4), 35–64.
URL: *https://www.aeaweb.org/articles?id=10.1257/mic.4.4.35*

Assenza, T., Heemeijer, P., Hommes, C. and Massaro, D. (2019), 'Managing self-organization of expectations through monetary policy: A macro experiment', *Journal of Monetary Economics* .
URL: *http://www.sciencedirect.com/science/article/pii/S0304393219302181*

Bao, T., Hennequin, M., Hommes, C. and Massaro, D. (2020), 'Coordination on bubbles in large-group asset pricing experiments', *Journal of Economic Dynamics and Control* **110**, 103702.
URL: *http://www.sciencedirect.com/science/article/pii/S0165188919300880*

Brock, W. A. and Hommes, C. H. (1997), 'A Rational Route to Randomness', *Econometrica* **65**(5), 1059–1095.
URL: *http://www.jstor.org/stable/2171879*

Clarida, R., Galí, J. and Gertler, M. (2000), 'Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory', *The Quarterly Journal of Economics* **115**(1), 147–180.
URL: *https://doi.org/10.1162/003355300554692*

Cornand, C. and Hubert, P. (2020), 'On the external validity of experimental inflation forecasts: A comparison with five categories of field expectations', *Journal of Economic Dynamics and Control* **110**, 103746.
URL: *http://www.sciencedirect.com/science/article/pii/S0165188919301459*

Cornea-Madeira, A., Hommes, C. and Massaro, D. (2019), 'Behavioral Heterogeneity in U.S. Inflation Dynamics', *Journal of Business & Economic*

Statistics **37**(2), 288–300.
URL: *https://ideas.repec.org/a/taf/jnlbes/v37y2019i2p288-300.html*

De Grauwe, P. (2010), 'Top-Down versus Bottom-Up Macroeconomics', *CESifo Economic Studies* **56**(4), 465–497.
URL: *https://doi.org/10.1093/cesifo/ifq014*

De Grauwe, P. (2011), 'Animal spirits and monetary policy', *Economic Theory* **47**(2-3), 423–457.
URL: *http://link.springer.com/10.1007/s00199-010-0543-0*

De Grauwe, P. and Ji, Y. (2020), 'Structural reforms, animal spirits, and monetary policies', *European Economic Review* **124**, 103395.
URL: *http://www.sciencedirect.com/science/article/pii/S0014292120300271*

Duffie, D. and Singleton, K. (1993), 'Simulated Moments Estimation of Markov Models of Asset Prices', *Econometrica* **61**(4), 929–52.
URL: *https://EconPapers.repec.org/RePEc:ecm:emetrp:v:61:y:1993:i:4:p:929-52*

Fagan, G., Henry, J. and Mestre, R. (2001), An area-wide model (AWM) for the euro area, Working Paper Series 42, European Central Bank.
URL: *https://ideas.repec.org/p/ecb/ecbwps/200142.html*

Franke, R. (2009), 'Applying the method of simulated moments to estimate a small agent-based asset pricing model', *Journal of Empirical Finance* **16**(5), 804 – 815.
URL: *http://www.sciencedirect.com/science/article/pii/S0927539809000425*

Franke, R. (2019), 'Bringing a Baseline Harrod-Kaldor Business Cycle Model to the Data. Second Draft'.

Franke, R., Jang, T.-S. and Sacht, S. (2015), 'Moment matching versus Bayesian estimation: Backward-looking behaviour in a New-Keynesian baseline model', *The North American Journal of Economics and Finance* **31**, 126 – 154.
URL: *http://www.sciencedirect.com/science/article/pii/S1062940814001156*

Gaunersdorfer, A., Hommes, C. H. and Wagener, F. O. (2008), 'Bifurcation routes to volatility clustering under evolutionary learning', *Journal of Economic Behavior & Organization* **67**(1), 27 – 47.
URL: *http://www.sciencedirect.com/science/article/pii/S0167268107001588*

Hommes, C. (2011), 'The heterogeneous expectations hypothesis: Some evidence from the lab', *Journal of Economic Dynamics and Control* **35**(1), 1 – 24.
URL: *http://www.sciencedirect.com/science/article/pii/S016518891000223X*

Hommes, C. and Lustenhouwer, J. (2019), 'Inflation targeting and liquidity traps under endogenous credibility', *Journal of Monetary Economics* **107**, 48 – 62.
URL: *http://www.sciencedirect.com/science/article/pii/S0304393219300170*

Hommes, C., Massaro, D. and Weber, M. (2019), 'Monetary policy under behavioral expectations: Theory and experiment', *European Economic Review* **118**, 193–212.
URL: *http://www.sciencedirect.com/science/article/pii/S0014292119300960*

Hommes, C., Sonnemans, J., Tuinstra, J. and van de Velden, H. (2005), 'Coordination of Expectations in Asset Pricing Experiments', *The Review of Financial Studies* **18**(3), 955–980.
URL: *https://doi.org/10.1093/rfs/hhi003*

Hommes, C., Sonnemans, J., Tuinstra, J. and van de Velden, H. (2008), 'Expectations and bubbles in asset pricing experiments', *Journal of Economic Behavior & Organization* **67**(1), 116 – 133.
URL: *http://www.sciencedirect.com/science/article/pii/S016726810700162X*

Jang, T.-S. and Sacht, S. (2016), 'Animal Spirits and the Business Cycle: Empirical Evidence from Moment Matching: Animal Spirits and the Business Cycle', *Metroeconomica* **67**(1), 76–113.
URL: *http://doi.wiley.com/10.1111/meca.12091*

Jang, T.-S. and Sacht, S. (2019), 'Forecast heuristics, consumer expectations, and New-Keynesian macroeconomics: A Horse race', *Journal of Economic Behavior & Organization* .
URL: *http://www.sciencedirect.com/science/article/pii/S0167268119300277*

Kukacka, J., Jang, T. and Sacht, S. (2018), 'On the Estimation of Behavioral Macroeconomic Models via Simulated Maximum Likelihood', *SSRN Electronic Journal* .
URL: *https://www.ssrn.com/abstract=3299450*

Kurz, M., Piccillo, G. and Wu, H. (2013), 'Modeling diverse expectations in an aggregated New Keynesian Model', *Journal of Economic Dynamics and Control* **37**(8), 1403 – 1433. Rethinking Economic Policies in a Landscape of Heterogeneous Agents.
URL: *http://www.sciencedirect.com/science/article/pii/S0165188913000390*

Lee, B.-S. and Ingram, B. (1991), 'Simulation estimation of time-series models', *Journal of Econometrics* **47**(2-3), 197–205.
URL: *https://EconPapers.repec.org/RePEc:eee:econom:v:47:y:1991:i:2-3:p:197-205*

Liu, C. and Minford, P. (2014), 'Comparing behavioural and rational expectations for the US post-war economy', *Economic Modelling* **43**, 407 – 415.
URL: *http://www.sciencedirect.com/science/article/pii/S026499931400340X*

McFadden, D. (1989), 'A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration', *Econometrica* **57**(5), 995–1026.
URL: *https://EconPapers.repec.org/RePEc:ecm:emetrp:v:57:y:1989:i:5:p:995-1026*

Milani, F. (2007), 'Expectations, learning and macroeconomic persistence', *Journal of Monetary Economics* **54**(7), 2065 – 2082.
URL: *http://www.sciencedirect.com/science/article/pii/S0304393206002406*

Pakes, A. and Pollard, D. (1989), 'Simulation and the Asymptotics of Optimization Estimators', *Econometrica* **57**(5), 1027–57.
URL: *https://EconPapers.repec.org/RePEc:ecm:emetrp:v:57:y:1989:i:5:p:1027-57*

Appendix A

A Note on Notation

The model described in Chapter 3 is a slightly modified version of the model from Hommes et al. (2019). This appendix describes these modifications and discusses its implications.

The forward-looking version of the baseline 3-equation NKM from Hommes et al. (2019) reads as follows:

$$y_t = \bar{y}^e_{t+1} - \varphi(i_t - \bar{\pi}^e_{t+1}) + g_t \tag{A.1}$$

$$\pi_t = \lambda y_t + \rho \bar{\pi}^e_{t+1} + u_t \tag{A.2}$$

$$i_t = \text{Max}\{\bar{\pi} + \phi_\pi(\pi_t - \bar{\pi}) + \phi_y(y_t - \bar{y}), 0\}, \tag{A.3}$$

where y_t is the output gap, π_t is inflation rate and r_t is the nominal interest rate. $\bar{y}^e_{t+1}$ and $\bar{\pi}^e_{t+1}$ are the expected output gap and inflation at time $t+1$, respectively. Both represent averages over population, and both are formed at time t when only information up to time $t-1$ is available. $\bar{\pi}$ and $\bar{y}$ are the steady-state inflation rate and output gap, respectively. g_t and u_t are random shocks.

Equation (A.3) is the central bank's monetary policy rule, equation (A.2) is the NKPC, and equation (A.1) is the dynamic IS curve. ϕ_π and ϕ_y are non-negative parameters which determine how the central bank reacts to the shocks to inflation rate and output gap, respectively. The shocks g_t and u_t are normally distributed random variables with zero mean and standard deviations σ_π and σ_y, respectively. φ and λ are non-negative parameters which represent the slopes of the IS curve and NKPC, respectively. The 'Max' operator in (A.3) represents zero lower bound for the nominal interest rate.

To prepare this system for further analysis and estimation, we make two

Table A.1: Mapping of the Notation

Original	New
φ	τ
g_t	ε_t
σ_y	σ_y
λ	κ
u_t	ε_t
σ_π	σ_π
ϕ_π	ϕ_π
ϕ_y	ϕ_y
η	ρ
β	γ
δ	η

adjustments. First, we set the central bank's inflation target $\bar{\pi}$ to 0. This implies zero steady-state output gap $\bar{y}$ (follows from equation 3.2) and zero steady-state nominal interest rate i_t (follows from equation 3.3). Removing zero lower bound on the nominal interest rate allows us to read and treat y_t, π_t, i_t as the output gap, inflation rate gap and interest rate gap, respectively.[14]

Second, in terms of symbols, we make use of the notation from Kukacka et al. (2018). This switch is summarized in Table A.1.

As a result, we arrive at the following 3-equation NKM:

$$y_t = \bar{y}^e_{t+1} - \tau(r_t - \bar{\pi}^e_{t+1}) + \varepsilon_{y,t} \tag{A.4}$$

$$\pi_t = \nu\bar{\pi}^e_{t+1} + \kappa y_t + \varepsilon_{\pi,t} \tag{A.5}$$

$$r_t = \phi_\pi \pi_t + \phi_y y_t, \tag{A.6}$$

which is exactly the system of equations described in Chapter 3.

Switching to this set of letters allows us to work with MATLAB scripts related to Kukacka et al. (2018), Jang and Sacht (2019), Jang and Sacht (2016) more conveniently; it keep the notation flexible for potential introduction of hybrid NKM into this work; and most importantly, it enables us to compare our estimation results to those of the authors above more directly.

It has been verified that the switch from non-zero steady state to zero steady state does not alter our conclusions, at least those related to Figure 3.2. In fact, since equations (3.1) and (3.2) "are typically derived by log-linearizing around a steady state with zero inflation rate" (Hommes et al. 2019), setting inflation target to zero might be a better, more general, choice. The reason why

[14] A more elaborated discussion of this treatment can be found in Franke et al. (2015).

inflation target is greater than zero in Hommes et al. (2019) is probably that these equations are used as the underlying model for their LtFE experiment; and since the goal of the participants of the experiment is to predict inflation rate as accurately as possible, it would be too easy for them to discover true inflation in their experiment in case it is set to 0%.

Appendix B

Methodology Figures

This appendix presents four figures closely related to the assessment of our estimation procedure from Chapter 4.

First, we present two sets of histograms. Figure B.1 shows how does the estimation procedure from Chapter 4 behaves when the pseudo-true DGP is HSM and we assume that it is RE – a byproduct of the assessment of the probability of type II error.

Similarly, Figure B.2 shows the distributions of the estimates when the roles are reversed – a byproduct of the assessment of the probability of type I error.

Figure B.1: Results of Simulations: HSM←RE

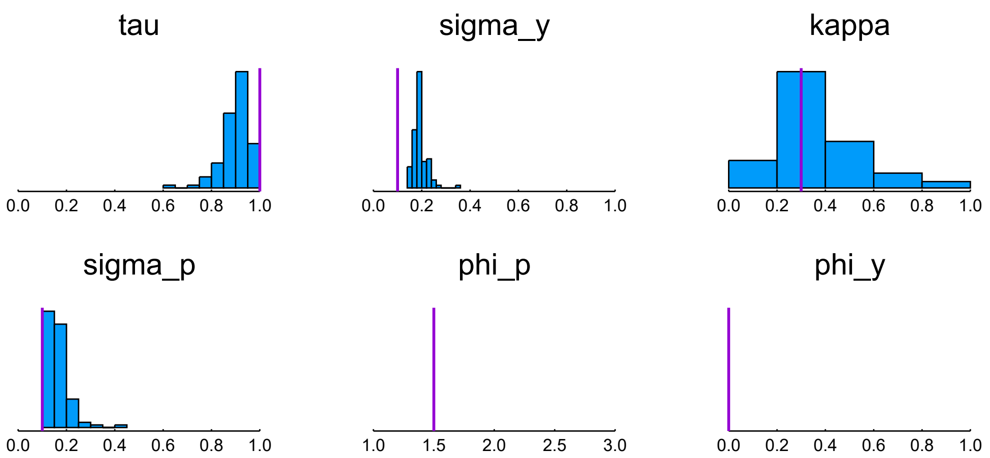
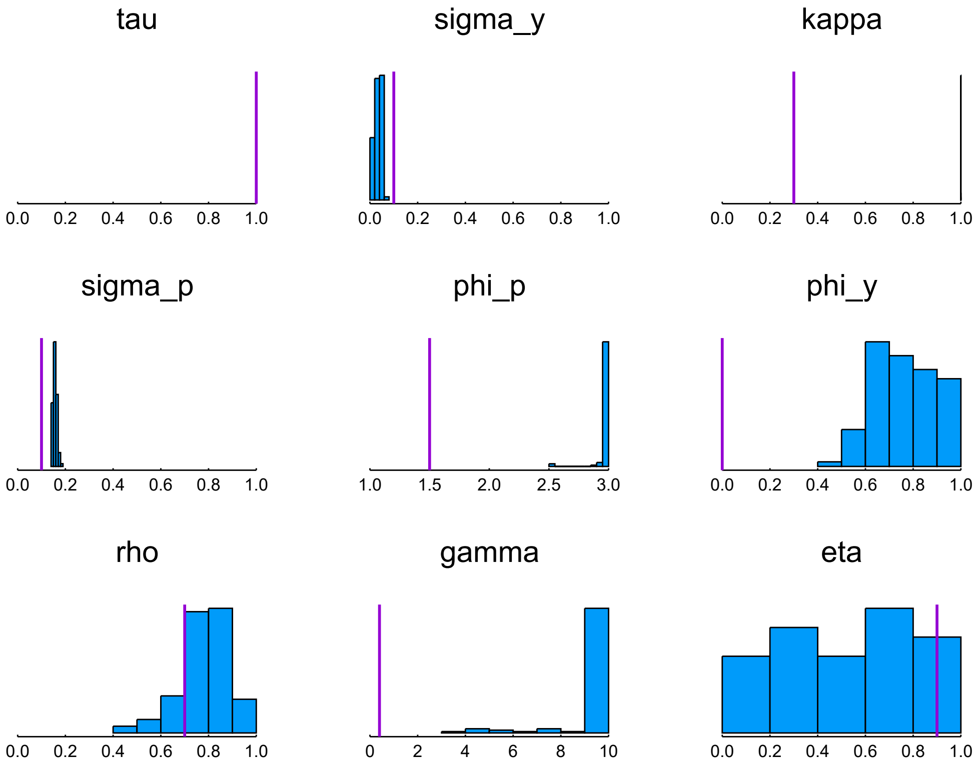


Figure B.2: Results of Simulations: RE←HSM



When reading the assessment of the probability of type I error in Chapter 4, one might wonder whether the differences between J^{HSM} and J^{RE} are stable across different random seeds. Figure B.3 depicts the distributions of both absolute and relative differences between J^{HSM} and J^{RE} when the pseudo-true model is RE.

Finally, Figure B.4 depicts the same for type II error, that is, the distributions of differences between J^{RE} and J^{HSM} when the pseudo-true model is HSM.

Figure B.3: Results of Simulations: Difference Between Fitness Measures for Type I Error

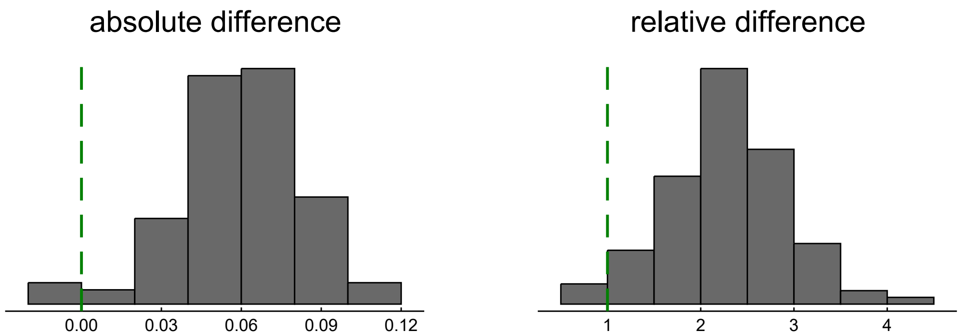
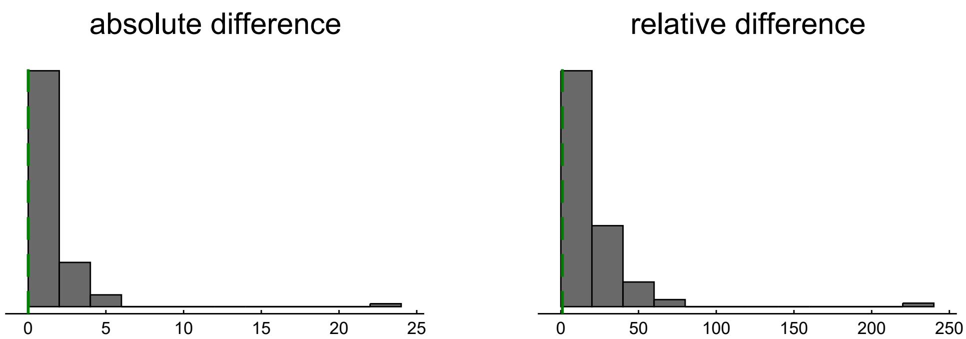


Figure B.4: Results of Simulations: Difference Between Fitness Measures for Type II Error



Appendix C

Results Figures

This appendix includes supplementary graphical materials for Chapter 5.

First, Figure C.1 depicts the macroeconomic data before the application of the Hodrick-Prescott filter.

Figure C.1: Empirical Time Series

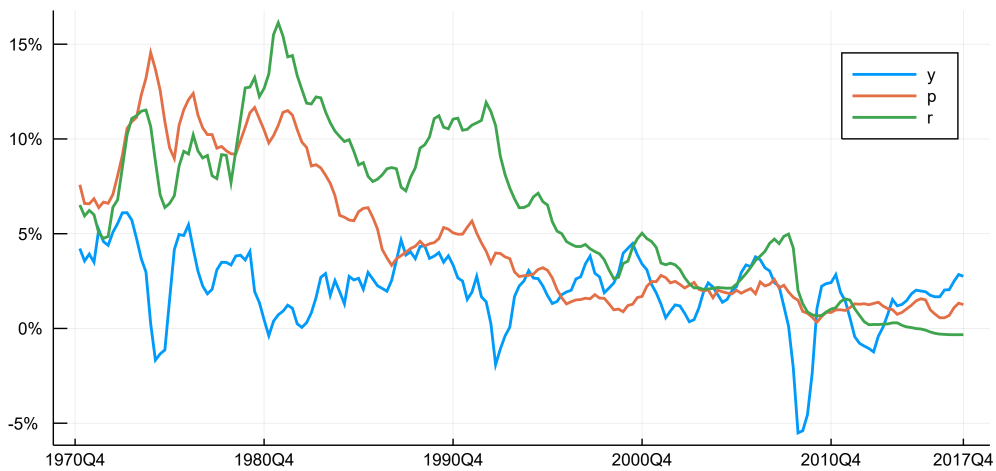
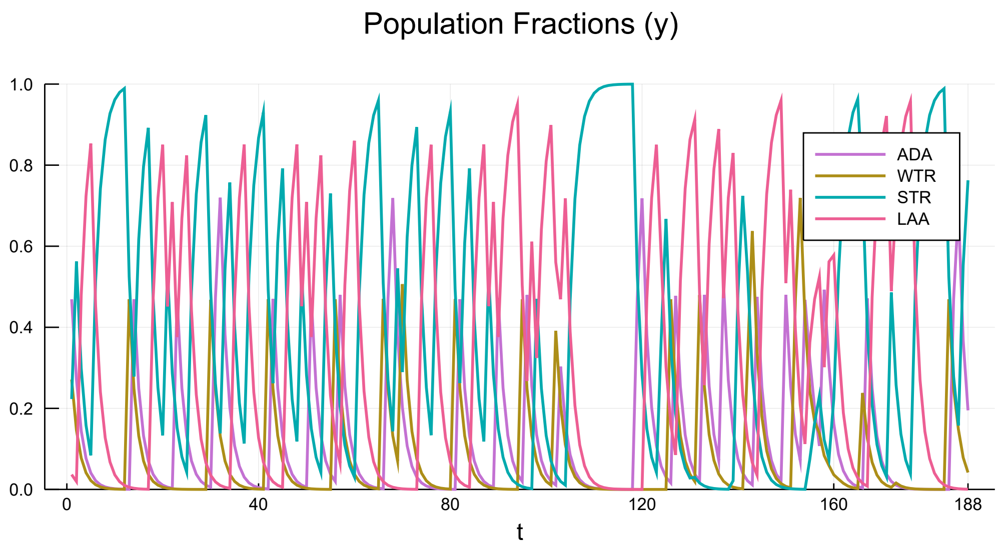


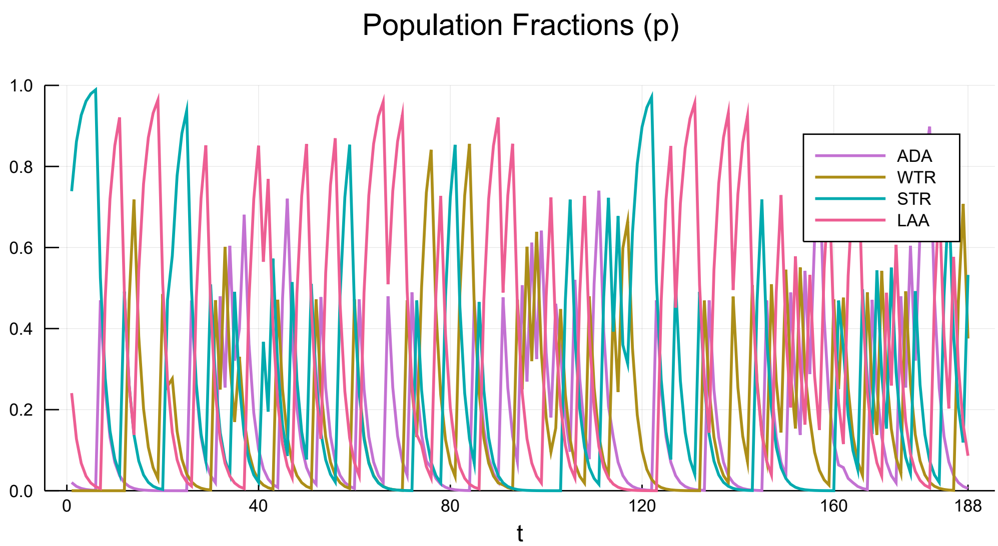
Table C.1: Results of the Estimation With a Binding Constraint

Parameter	HSM	RE
τ	0.35	0.04
<0, 1>		
σ_y	0.28	1.00
<0, 1>		
κ	0.04	0.00
<0, 1>		
σ_π	0.03	0.69
<0, 1>		
ϕ_π	2.15	1.58
<1, 3>		
ϕ_y	0.46	0.22
<0, 1>		
ρ	0.04	-
<0, 1>		
γ	9.69	-
<0, 10>		
η	0.53	-
<0, 1>		
J	6.45	137.16

Next, Table C.1 presents the results of the very first estimation of the models, when the upper bound of the search range for σ_y is binding.

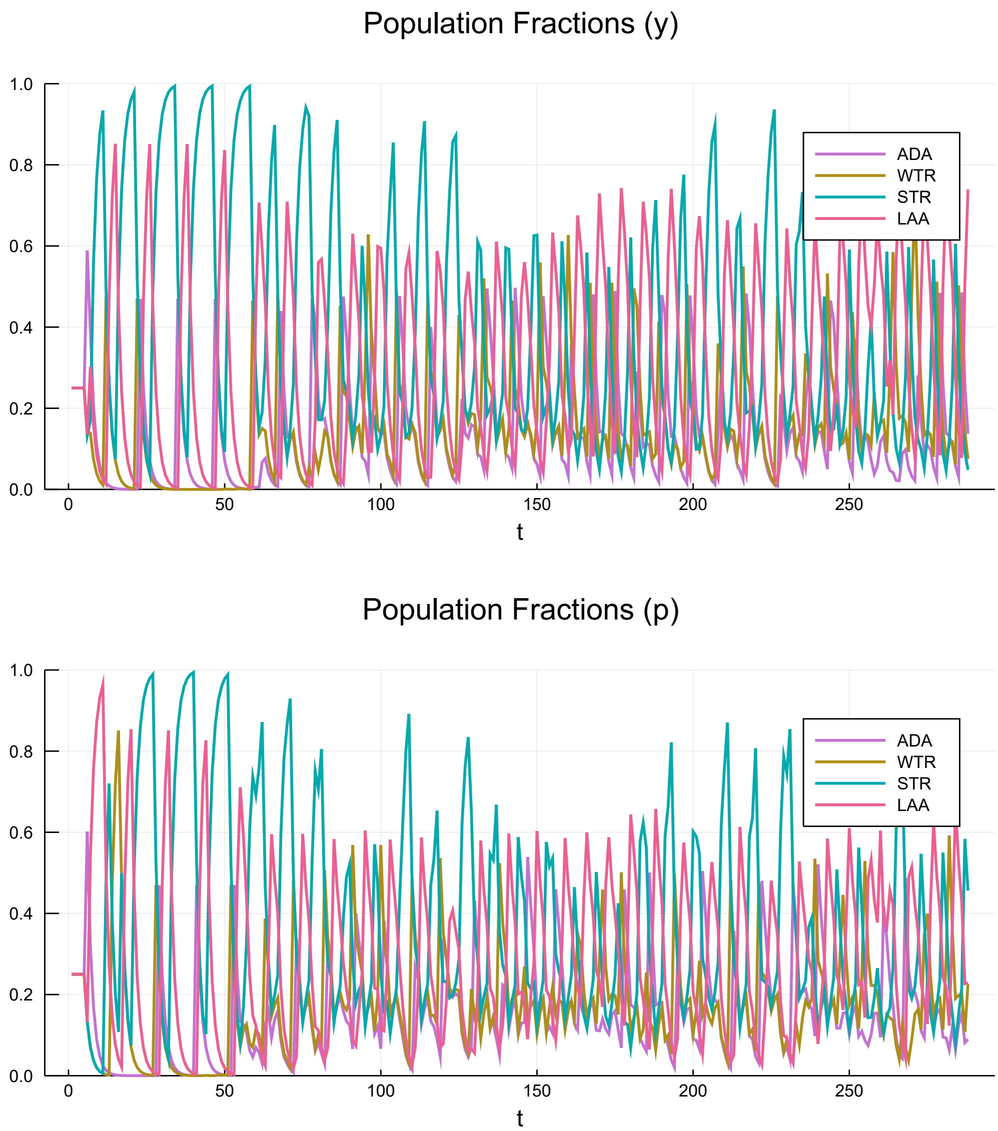
Figure C.2: Results of Estimation: HSM Population Fractions





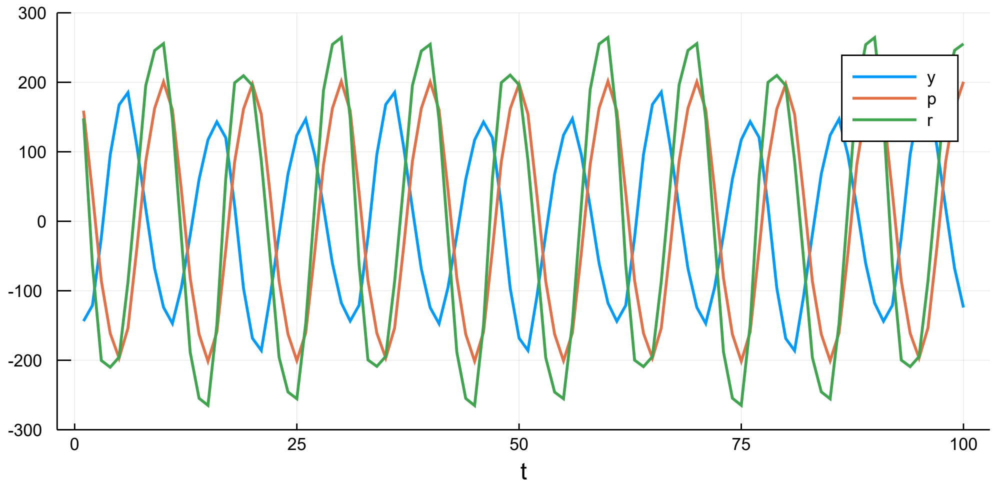
As a complement to Figure 5.2, Figure C.2 shows the underlying population fractions for HSM.

Figure C.3: An Example of HSM Explosion ($\phi_y = 0.00$) – Population Fractions



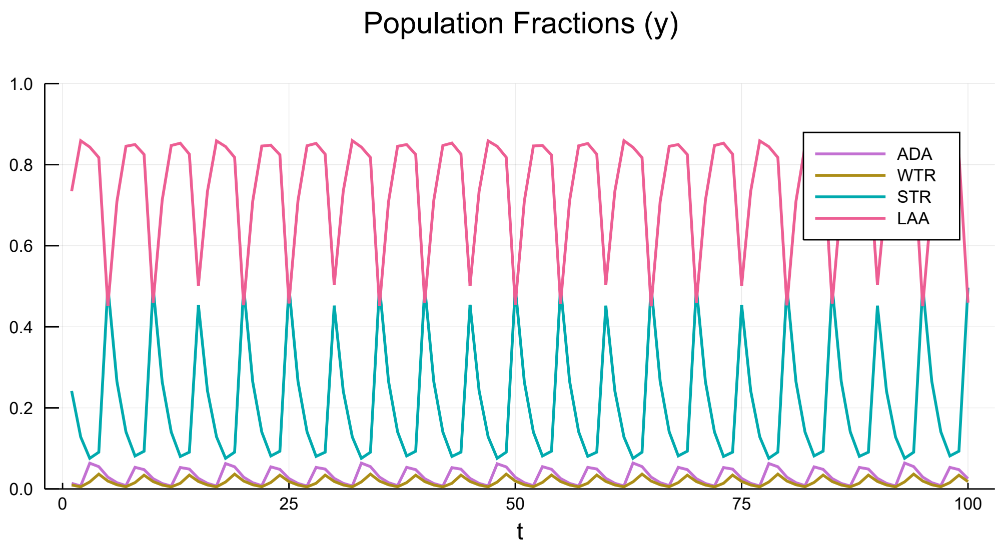
Related to the analysis illustrated in Figure 5.3, Figure C.3 shows the development of population fractions over time for the realization with $\phi_y = 0.00$.

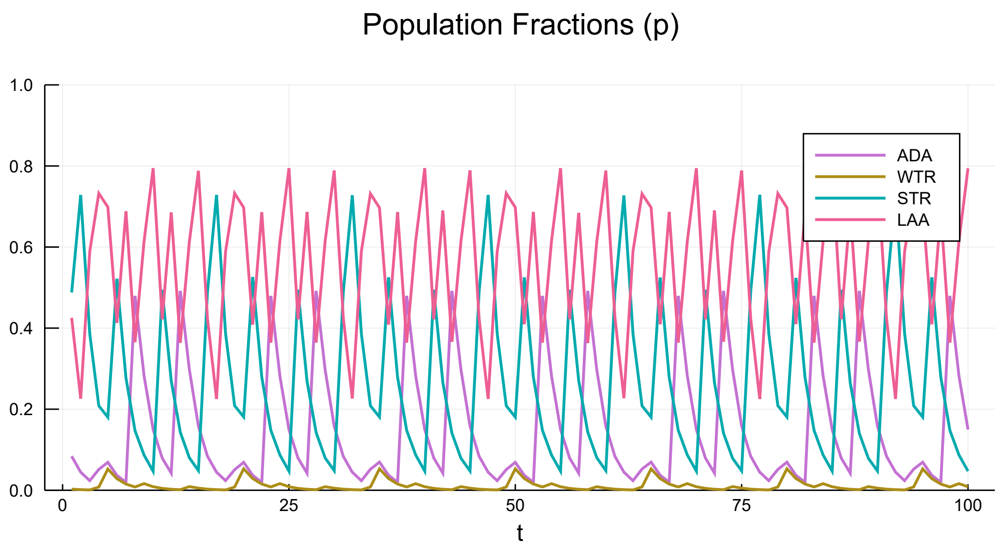
Figure C.4: An Example of HSM Oscillation ($\phi_y = 1.20$) – a Deeper Look



When looking at Figure 5.6, one might be curious how exactly do the macroeconomic variables oscillate. Figure C.4 provides an answer.

Figure C.5: An Example of HSM Oscillation ($\phi_y = 1.20$) – a Deeper Look at the Population Fractions





Finally, Figure C.5 shows the population fractions counterpart of the above.