

Kukacka, Jiri; Jang, Tae-Seok; Sacht, Stephen

Working Paper

On the estimation of behavioral macroeconomic models via simulated maximum likelihood

Economics Working Paper, No. 2018-11

Provided in Cooperation with:

Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Kukacka, Jiri; Jang, Tae-Seok; Sacht, Stephen (2018) : On the estimation of behavioral macroeconomic models via simulated maximum likelihood, Economics Working Paper, No. 2018-11, Kiel University, Department of Economics, Kiel

This Version is available at:

<https://hdl.handle.net/10419/190330>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

C | A | U

Christian-Albrechts-Universität zu Kiel

Department of Economics

Economics Working Paper
No 2018-11

On the Estimation of Behavioral Macroeconomic Models via Simulated Maximum Likelihood

by Jiri Kukacka, Tae-Seok Jang, Stephen Sacht

issn 2193-2476



On the Estimation of Behavioral Macroeconomic Models via Simulated Maximum Likelihood

December 8, 2018

Jiri Kukacka^a, Tae-Seok Jang^{b,*}, Stephen Sacht^c

^a*Institute of Economic Studies, Faculty of Social Sciences, Charles University, Opletalova 26, 110 00 Prague 1, Czechia*

^b*School of Economics & Trade, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 41566, Republic of Korea*

^c*Department of Economics, Kiel University, Olshausenstrasse 40, 24118 Kiel, Germany*

Abstract

In this paper, we introduce the simulated maximum likelihood method for identifying behavioral heuristics of heterogeneous agents in the baseline three-equation New Keynesian model. The method is extended to multivariate macroeconomic optimization problems, and the estimation procedure is applied to empirical data sets. This approach considerably relaxes restrictive theoretical assumptions and enables a novel estimation of the intensity of choice parameter in discrete choice. In Monte Carlo simulations, we analyze the properties and behavior of the estimation method, which provides important information on the behavioral parameters of the New Keynesian model. However, the curse of dimensionality arises via a consistent downward bias for idiosyncratic shocks. Our empirical results show that the forward-looking version of both the behavioral and the rational model specifications exhibits good performance. We identify potential sources of misspecification for the hybrid version. A novel feature of our analysis is that we pin down the switching parameter for the intensity of choice for the Euro Area and US economy.

JEL: C53, D83, E12, E32

Keywords: Behavioral Heuristics; Intensity of Choice; Monte Carlo Simulations; New-Keynesian Model; Simulated Maximum Likelihood

*Corresponding author

Email addresses: jiri.kukacka@fsv.cuni.cz (Jiri Kukacka), taeseok.jang@gmail.com (Tae-Seok Jang), sacht@economics.uni-kiel.de (Stephen Sacht)

1. Introduction

Since the outbreak of the global financial crisis at the end of the last decade, which resulted in a contagion effect on real economic activity and in turn low economic growth, formerly axiomatic beliefs in the rational expectation paradigm have eroded. Macroeconomists are now fully aware of the crucial effect of non-rational expectation formation on the future development of key economic variables. As a result, various approaches advocate the design of optimal policy strategies within a behavioral model framework, while the alternative explanation of expectation formation has rapidly gained attention in macroeconomics in recent years. That is in large part due to both existing suspicions regarding the rational expectation paradigm and the availability of powerful computers that enable extensive numerical simulations of macroeconomic models (Hommes, 2006; Windrum et al., 2007; Fagiolo et al., 2007; LeBaron and Tesfatsion, 2008; De Grauwe, 2011, 2012; Deák et al., 2017; Gabaix, 2017).

The decision to allow for heterogeneity in different groups of agents, who form their individual expectations in a non-rational way, has opened a new debate along both the theoretical and empirical dimensions. Related to this development, we now face the question of how monetary and fiscal policy interventions can be conducted if agents do not necessarily forecast economic developments given full information on the underlying structure of the economy and properties of the given exogenous shocks. Under this regime, expectation formation is achieved via helpful task-general procedures such as heuristics. The latter represent rules of behavior that stem from the fact that the structure of the economy is observable but the interactions of relevant variables such as output and inflation are barely comprehensible (Munier et al., 1999). Such boundedly rational behavior in response to a lack of all the information needed for proper forecasting is instead focused on habits, imitation and/or procedural optimization (Day and Pingle, 1991). In the absence of convenient analytic tractability with respect to dynamic stochastic general equilibrium (DSGE) models under rational expectations, macroeconomists instead face the challenge of non-linear modeling strategies under bounded rationality.

As heuristics represent specific forecasting rules, agents sort themselves into the corresponding group populated by all individuals who believe in this particular mechanism of expectation formation. The recently dominant mechanism for sorting (or, better, switching) applied in DSGE models is the discrete choice approach. As a result, from period to period, endogenous waves of economic beliefs such as optimism and pessimism are generated and lead to fluctuations in key economic variables, even in the absence of autocorrelated exogenous shocks.

While the discrete choice approach has become a prominent feature of models in financial economics (see Brock and Hommes, 1997, or Gaunersdorfer et al., 2008, among others) over the last two decades, it has also become a prominent candidate for inclusion in macroeconomic system-of-equations models such as the DSGE in recent years. We recognize De Grauwe (2011)'s work as one of the most significant contributions to behavioral macroeconomics. Indeed, since in this paper we are concerned with the estimation of a business cycle model to analyze short-run fluctuations, we focus on the discrete choice approach in the spirit of Jang and Sacht (2016), among others. In addition, studies from the field of experimental economics reveal that switching based on discrete choice can indeed be observed within a laboratory environment. We refer to Assenza et al. (2014) for a literature review that includes reports on the corresponding evidence in a DSGE context.

However, taking various forms of DSGE models to the empirical data is a daunting and complicated task. For example, endogenous macroeconomic variables in DSGE models are strongly influenced by stochastic demand and supply shocks if the notion of perfect foresight is abandoned as a mechanism to control the expectation formation process. Although a stochastic system can be represented in the state-space model, analytic solutions of non-linear models seldom exist for estimation. Further, these economic systems are sensitive to history: a small change in the noise will cause chaotic changes in the system trajectory. Hence, the joint probability density of economic variables is often analytically intractable, and we cannot easily apply likelihood inference for the parameters, because it is not possible to integrate the function over all parameters. Indeed, DSGE models can be expressed by the measurement equation only in the state-space representation, because the macroeconomic variables do not have a direct effect on the state recursion itself. However, agent-based macroeconomic models include the state of the economy based on agents' switching process, and the transition of economic states emphasizes the role of expectation formation in macroeconomic dynamics.

These challenges to estimation have motivated the use of simulation-based inference to address the statistical inference on intractable model dynamics. The recent advent of increased operational capacity of computers that allow for the processing of larger simulation-based macroeconomic has permitted developments in macroeconomic modeling. This makes it possible to obtain statistical inference for parameter estimation without knowing the analytic form of the model solutions. For example, the moment conditions are replaced by simulated counterparts, and more accurate estimation can be obtained through long simulations of behavioral variables (Franke, 2009; Franke and Westerhoff, 2012; Jang and Sacht, 2016; Grazzini and Richiardi, 2015; Franke, 2018). In particular, if prior distributions of macroeconomic variables are widely accepted, then we can rely on Bayesian inference in parameter estimation (Grazzini et al., 2017; Deák et al., 2017). The Bayesian approach is particularly relevant for the estimation of macroeconomic models because the sample size is often not large enough to ensure asymptotic statistical properties. This approach provides a means of resolving identification problems for large-scale DSGE models. An interesting comparison between Bayesian estimation and moment matching is provided by Franke et al. (2015). Another possible approach to estimating models with (sometimes latent) behavioral variables is based on recent advances in the Markov Chain Monte Carlo estimation technique (Doucet et al., 2001). In sequential Monte Carlo, also known as a particle filter or bootstrap filter, the distribution of hidden variable(s) can be approximated by a swarm of particles representing discrete draws from this distribution that are propagated from one time step to the next via a sampling-importance resampling algorithm (Lux, 2018). The growing interest in macroeconomic agent-based models has recently been accompanied by novel approaches for their empirical validation (Guerini and Moneta, 2017; Barde and van der Hoog, 2017; Lamperti, 2018; Lamperti et al., 2018). Lux and Zwinkels (2018) provide an excellent survey of these attempts.

This paper follows the contributions of Kristensen and Shin (2012), Lee and Song (2015), and Kukacka and Barunik (2017) where the conditional density of each observation is approximated via a standard kernel method while the approximation error is small enough to permit inference on the true model parameters and the time development of behavioral changes. Derivation of the simulated maximum likelihood estimator (SMLE) is then similar to that of the maximum likelihood estimator (MLE). Via the kernel approximation, the method elegantly bypasses restrictive

distributional assumptions of the MLE and the Bayesian estimation technique and the moment selection problem of the simulated method of moments (SMM). The contribution of our approach is to extend the estimation method from univariate financial applications to multivariate macroeconomic optimization problems and the first application of the multivariate SMLE framework for the estimation of a New-Keynesian model (NKM).

The most important parameter in models that consider a discrete choice switching mechanism is doubtless the parameter that measures agents' intensity of choice. This parameter is bounded from below with limit zero and defined over an open range up to infinity. An increase in the intensity of choice implies an increase in agents' sensitivity regarding their performance based on a fitness measure. The value of the parameter is therefore crucial to the degree of fluctuation in the economic indicators with respect to the switching process from one group to another. In this paper, we successfully pin down the empirical value of the intensity of choice for the Euro Area and the US economy. Several previous studies in economics and finance appear to have had difficulty estimating this model parameter, as a behavioral model with discrete choice exhibits a much more complex non-linear structure than a linearized DSGE model under rational expectations.

While the estimation of behavioral macroeconomic models is relatively new, the framework of choice is based upon the three-equation NKM structure, which relies on the dynamic IS and Phillips curves and a Taylor rule. [Milani \(2007\)](#) represents the first prominent example of the estimation of this standard workhorse model in its behavioral representation. By applying Bayesian techniques, the author empirically demonstrates that backward-looking components resulting from habit formation and price indexation can be neglected; a constant-gain learning approach is introduced to describe the persistence in the data well. [Liu and Minford \(2014\)](#) estimate the purely forward-looking version of the standard NKM via the indirect inference approach. They find a low value for the intensity of choice parameter but also a low degree of variation. This is at odds with the premise that calibrated values have to be substantially at variance with the data's requirements. The authors claim that the reason for this observation stems from the underlying behavioral expectation formation approach. [Grazzini et al. \(2017\)](#) estimate a hybrid version of the same model with leads and lags via Bayesian techniques. They report uninformative estimates for the degree of switching, where the posterior mean is centered around its pseudo-true value but the corresponding posterior distribution does not depart significantly from the prior one. [Deák et al. \(2017\)](#) also use Bayesian estimation to compare forecasting errors under rational expectations (RE) with those under bounded rational (BR) and learning in a New Keynesian behavioral model. [Jang and Sacht \(2016\)](#) and [Jang and Sacht \(2018\)](#) consider both the purely forward-looking and hybrid version of the NKM. The authors perform a calibration and sensitivity exercise with respect to this specific parameter of interest and apply the simulated method of moments approach. This appears to be a reasonable approach since the authors report insignificant parameter estimates otherwise while the fitness of the models dropped significantly when the parameter is estimated freely. For a more elaborated survey on the estimation of recent macroeconomic behavioral models under discrete choice, we refer to [Franke and Westerhoff \(2018\)](#).

As a novelty, in our study we are able to report significant parameter estimates for the intensity of choice. This holds across all data sets for the two different economic regions we consider. While for the Euro Area we rely on data from 1970 to 2009 only due to limited availability, in case of the

US economy we split all observations into three different data sets. Therefore, we account for the changes in monetary policy over time during the Pre-Volcker, Volcker & Greenspan and Bernanke & Yellen eras as chairs of the Federal Reserve. The corresponding monetary policy regimes are denoted commonly as the Great Inflation (1960 until 1979), the Great Moderation (1980 until 2009) and the Great Recession (2010 until 2016) periods. This differentiation is well-known and necessary, as we observe high and low degrees of volatility in inflation during the Great Inflation and Moderation periods, respectively, and a nominal interest rate at the zero lower bound during the Great Recession. Additionally, we find that the switching process in heuristics is more intense in the Euro Area than the US economy for both the forward-looking and hybrid versions of the New Keynesian model. The results in our paper thus help to understand the implications of monetary policy in different regimes in the absence of rational expectations.

The paper proceeds as follows. In [Section 2](#), we explain the baseline New-Keynesian model with heterogeneous agents. In [Section 3](#), we discuss the simulated maximum likelihood estimator method that is adapted for behavioral macroeconomic models. In [Section 4](#), we conduct Monte Carlo simulations for the properties of the estimation method before we discuss empirical estimation results in [Section 5](#). Finally, in [Section 6](#), we conclude the paper. Technical details and additional results are relegated to the Appendix.

2. Expectation Formation in the Baseline NKM

The baseline NKM reads as follows:

$$y_t = \frac{1}{1+\chi} \tilde{E}_t^j y_{t+1} + \frac{\chi}{1+\chi} y_{t-1} - \tau(r_t - \tilde{E}_t^j \pi_{t+1}) + \varepsilon_{y,t}, \quad (1)$$

$$\pi_t = \frac{\nu}{1+\alpha\nu} \tilde{E}_t^j \pi_{t+1} + \frac{\alpha}{1+\alpha\nu} \pi_{t-1} + \kappa y_t + \varepsilon_{\pi,t}, \quad (2)$$

$$r_t = \phi_r r_{t-1} + (1 - \phi_r)(\phi_\pi \pi_t + \phi_y y_t) + \varepsilon_{r,t}, \quad (3)$$

where the superscript $j = \{\text{RE}, \text{BR}\}$ refers to the RE and the BR model, respectively. The corresponding expectations operator is \tilde{E}_t^j , which has to be specified for both models. All variables are given in gap notation, i.e., by the difference of their levels from the underlying steady states. All variables are given in quarterly magnitudes.

In equation (1), the dynamic IS curve results from intertemporal optimization of consumption and saving, which leads to consumption smoothing. The parameter $\tau \geq 0$ denotes the inverse intertemporal elasticity of substitution in consumption behavior. Equation (2) represents the New-Keynesian Phillips Curve (NKPC), where the output gap (y_t) acts as the driving force behind inflation dynamics from monopolistic competition and Calvo-type sticky prices. The slope of the NKPC is given by the parameter $\kappa \geq 0$. ν measures the discount factor ($0 < \nu < 1$). Hybridity is incorporated into the demand and supply framework using the parameters for habit formation $0 \leq \chi \leq 1$ and price indexation $0 \leq \alpha \leq 1$, respectively. According to the Taylor rule (3) with interest rate smoothing ($\phi_r \geq 0$), the monetary authority reacts directly to contemporaneous movements in output ($\phi_y \geq 0$) and inflation ($\phi_\pi \geq 0$). We assume that the exogenous driving forces in the model variables follow idiosyncratic shocks $\varepsilon_{s,t}$, which are independent and identically distributed around mean zero and variance σ_s^2 with variables $s = \{y, \pi, r\}$.

Under RE, the forward-looking terms are described by the expectations of the output gap and inflation gap at time $t + 1$ in equations (1) and (2):

$$\tilde{E}_t^{RE} z_{t+1} = E_t z_{t+1} + E_t \tilde{\varepsilon}_{z,t}, \quad (4)$$

with $z = \{y, \pi\}$, and where E_t denotes the statistical expectation operator conditional on information at time t . For the random error term, which is independent of the future realizations in z , $E_t(\tilde{\varepsilon}_{z,t}) = 0$ holds. This implies that agents' expectations are not systematically biased under rational expectations.

For the BR specification of the model, we distinguish between expectation formation separately with respect to the output and inflation gap. Regarding the output gap expectations, agents are able to sort themselves into two groups of forecasters expressed through the following heuristics:

$$E_t^{OPT} y_{t+1} = \frac{1}{2} \cdot [\beta + \delta \lambda_{y,t}], \quad (5)$$

$$E_t^{PES} y_{t+1} = -\frac{1}{2} \cdot [\beta + \delta \lambda_{y,t}]. \quad (6)$$

For equations (5) and (6), we apply the heuristics imposed by De Grauwe (2011) when modeling the divergence in beliefs. Here, we assume that agents may adopt either an *optimistic* or a *pessimistic* attitude towards movements in the future output gap (in the following indicated by the superscripts *OPT* and *PES*, respectively). Hence, boundedly rational agents are uncertain about the future dynamics of the output gap and therefore predict a subjective mean value of y_{t+1} measured by $\beta \geq 0$. However, this kind of subjective forecast is generally biased and therefore depends on the volatility of the output gap, i.e., given by the time-dependent unconditional standard deviation $\lambda_{y,t} \geq 0$. The corresponding parameter $\delta \geq 0$ measures the degree of divergence in the movement of economic activity. We consider symmetry with respect to β and δ : optimists expect that the output gap will differ positively from the steady state value \bar{y} (which is equal to zero by assumption), while pessimists will expect a negative deviation of the same magnitude.

For the BR specification, the switching from one group to the other is based on discrete choice theory and is described in general as follows. The expression for the market forecast regarding the output gap across the two groups is given by

$$\tilde{E}_t^{BR} y_{t+1} = \sum_{i=1}^2 (\alpha_{y,t}^{k\{i\}} \cdot E_t^{k\{i\}} y_{t+1}), \quad (7)$$

with $k = \{OPT, PES\}$. The probability $\alpha_{y,t}^k$ indicates stochastic behavior by the agents who adopt a particular forecasting rule, i.e., based on equation (5) or (6). Then, $\alpha_{y,t}^k$ can be interpreted as the probability of being an optimist or a pessimist with respect to development of the output gap in period t . The selection of the forecasting rules (5) and (6) depends on the forecast performance of each group U_t^k , which is given by the mean squared forecasting error. The utility of forecast performance can be simply updated in every period as (cf. Brock and Hommes, 1997):

$$U_{y,t}^k = \rho U_{t-1}^k - (1 - \rho)(E_{t-2}^k y_{t-1} - y_{t-1})^2, \quad (8)$$

where the parameter ρ is used to measure symmetrically the memory of the two different types of agents ($0 \leq \rho \leq 1$). Here, $\rho = 0$ suggests that agents have no memory of past observations, while $\rho = 1$ means that they have infinite memory instead. Agents can revise their expectations by applying discrete choice theory given forecast performance. The different types of performance measures can be utilized for $\alpha_{y,t}^k$ as follows:

$$\alpha_{y,t}^{OPT} = \frac{\exp(\gamma U_t^{OPT})}{\sum_{i=1}^2 \exp(\gamma U_t^{k\{i\}})}. \quad (9)$$

The parameter $\gamma \geq 0$ denotes the intensity of choice. The latter is crucial to the stability of the system (cf. [Hommes, 2001](#), as well as [Jang and Sacht, 2016](#), among others). Of course, the probability of being a pessimist is then simply given by

$$\alpha_{y,t}^{PES} = 1 - \alpha_{y,t}^{OPT}. \quad (10)$$

According to [De Grauwe \(2011\)](#), the central bank seeks to control for the value of the inflation gap via the interest channel of monetary policy. Therefore, it anchors expectations by announcing a target for the inflation gap given by $\bar{\pi}$. Without loss of generality, $\bar{\pi} = 0$ holds. The group of so-called *inflation targeters* (denoted by the superscript *TAR*) consider this pre-commitment strategy to be fully credible. The corresponding forecasting rule then becomes

$$E_t^{TAR} \pi_{t+1} = \bar{\pi}. \quad (11)$$

Conversely, the group of *inflation extrapolators* (denoted by the superscript *EXP*) will expect that the future value of the inflation gap is simply given by

$$E_t^{EXT} \pi_{t+1} = \pi_{t-1}. \quad (12)$$

Clearly, equations (7) to (10) have to be adjusted for the inflation gap expectation formation process. Note that the memory parameter given by ρ is equal for both the output and inflation gaps.

The state space representation of the NKM is given by:

$$\mathbf{A}\mathbf{X}_t + \mathbf{B}\mathbf{X}_{t+1}^j + \mathbf{C}\mathbf{X}_{t-1}^j + \mathbf{D}\mathbf{\Gamma}_t = \mathbf{0}, \quad (13)$$

with $\mathbf{X}_t = (y_t, \pi_t, r_t)'$, $\mathbf{X}_{t+1}^j = (\tilde{E}_t^j y_{t+1}, \tilde{E}_t^j \pi_{t+1}, \tilde{E}_t^j r_{t+1})'$, $\mathbf{X}_{t-1}^j = (y_{t-1}, \pi_{t-1}, r_{t-1})'$ and $\mathbf{\Gamma}_t = (\varepsilon_{y,t}, \varepsilon_{r,t}, \varepsilon_{\pi,t})'$. The corresponding general reduced-form solution of

$$\mathbf{X}_t = -\mathbf{A}^{-1}[\mathbf{B}\mathbf{X}_{t+1}^j + \mathbf{C}\mathbf{X}_{t-1}^j + \mathbf{D}\mathbf{\Gamma}_t] \quad (14)$$

is then obtained by applying the method of undetermined coefficients. In the BR case, the forward-looking elements in X_{t+1}^{BR} are replaced by the forecast heuristics (5) and (6), obviously except for $\tilde{E}_t^j r_{t+1}$. It follows that in this case, this non-linear model specification has to be solved by backward induction.

Following the approach developed by [Jang and Sacht \(2016\)](#), we estimate the purely forward-looking and hybrid versions of the model and consider two different specifications of the expectation forma-

tion process. Regarding the first specification, the both model versions are estimated based on RE. The second specification reflects BR expectations, where equations (5) and (6) as well as (11) and (12) are applied. In particular, the latter specification accounts for the impact of the well-known animal spirits paradigm, i.e., waves of optimistic and pessimistic beliefs occur. Therefore, we apply the SMLE method, which is primarily known from the financial econometrics literature.

For the purpose of introducing the SMLE approach into the macroeconomic literature, in the first step, we focus on the purely forward-looking version of the NKM only, which implies $\chi = \alpha = \phi_r = 0$ in the baseline NKM. Therefore, in the following section, we describe how we estimate the *structural* ($\tau, \kappa, \phi_\pi, \phi_y$) and the *bounded rational* (β, δ, γ) parameters and the corresponding *standard deviations* ($\sigma_y, \sigma_\pi, \sigma_r$) of the model specifications. When we turn to our empirical study in section [Section 5](#), we consider both versions of the model. In the hybrid case, we estimate the parameters for habit formation, price indexation, and interest rate smoothing, χ , α , and ϕ_r , respectively, in conjunction with the remaining ones obtained according to the system of equations (1) to (3).

3. The Simulated Maximum Likelihood Approach for the NKM

This section extends the SMLE method, which is primarily known from the (financial) econometric literature ([Kristensen and Shin, 2012](#); [Lee and Song, 2015](#); [Kukacka and Barunik, 2017](#)), to a macroeconomic estimation problem. Let us assume a generic multivariate time series process (z_t, x_t) , $z_t : t \mapsto \mathbb{R}^l$, $l \in \mathbb{N}$; $x_t : t \mapsto \mathcal{X}_t$, $t = 1, \dots, \infty$. Suppose that we have T realizations $\{(z_t, x_t)\}_{t=1}^T$. We further assume that the time series $\{z_t\}_{t=1}^T$ has been generated by a fully parametric model:

$$z_t = m_t(x_t, \varepsilon_t, \theta), \quad t = 1, \dots, T, \quad (15)$$

where a model function $m : \{x_t, \varepsilon_t, \theta\} \mapsto \mathbb{R}^l$, $\theta \in \Theta \subseteq \mathbb{R}^n$ is an unknown parameter vector, and $\varepsilon_t \in \mathbb{R}^l$ is an independent and identically distributed (i.i.d.) sequence with known distribution \mathcal{F}_ε , which is assumed to be disconnected from t or θ . In general, both multivariate processes (z_t, x_t) can be non-stationary, realizations z_t can directly represent empirical observables or can be latent variables, and x_t is assumed to contain exogenous explanatory variables, including dependent variables z_t and their lags, or even more generally, the space \mathcal{X}_t can be time-varying. Finally, we assume the model to have an associated conditional density $p_t(z|x;\theta)$:

$$P(z_t \in A | x_t = x) = \int_A p_t(z|x;\theta) dz, \quad t = 1, \dots, T, \quad (16)$$

for any Borel set $A \subseteq \mathbb{R}^l$.

Let us now consider the case of the purely forward-looking version of the NKM suggested in [Section 2](#). Here, $z_t = \{y_t, \pi_t, r_t\}$, $l = 3$, from equations (1) to (3); x_t contains dependent variables z_t and their lags plus other terms collected from equations (1) to (12) that together fully represent the model function m_t , i.e., the discount factor ν , window size for a moving average computation of the time-dependent unconditional standard deviation of the output gap $\lambda_{y,t}$, agents' memory ρ , and the target for the inflation gap $\bar{\pi} = 0$. ε_t generally stands for a set of idiosyncratic shocks $\{\varepsilon_{y,t}, \varepsilon_{\pi,t}, \varepsilon_{r,t}\}$ from equations (1) to (3), which are i.i.d. around mean zero and variance σ_s^2 with variables $s = \{y, \pi, r\}$. Finally, for the RE NKM $\theta = \{\tau, \kappa, \phi_\pi, \phi_y, \sigma_y, \sigma_\pi, \sigma_r\}$, i.e., $n = 7$, while for

the BR NKM $\theta = \{\tau, \kappa, \phi_\pi, \phi_y, \beta, \delta, \gamma, \sigma_y, \sigma_\pi, \sigma_r\}$, i.e. $n = 10$. Note that $\chi = \alpha = \phi_r = 0$ holds.

As a result of adaptive revisions of agents' expectations introduced by equation (9), the probability $p_t(z|x; \theta)$ from equation (16) does not have a closed-form representation. Therefore, an exact mathematical derivation of the likelihood function of the model in equation (15) does not exist, and a standard estimator of θ , the maximizer of the conditional log-likelihoods

$$\tilde{\theta} = \arg \underbrace{\max}_{\theta \in \Theta} L_T(\theta), \quad (17)$$

where $L_T(\theta) = \sum_{t=1}^T \log p_t(z_t|x_t; \theta)$, is infeasible.¹

However, we are always able to numerically obtain simulated observations from the model in equation (15). The SMLE method presented below then allows us to compute a simulated conditional density, which we use to obtain a simulated version of the MLE.

To obtain a simulated approximation of the conditional density $p_t(z_t|x_t; \theta) \forall t \in \langle 1, \dots, T \rangle, z_t \in \mathbb{R}^l, x_t \in \mathcal{X}_t$, and $\theta \in \Theta$, we first generate $N \in \mathbb{N}$ i.i.d. draws from the l -dimensional distribution $\mathcal{F}_\varepsilon, \{\varepsilon_i\}_{i=1}^N$, to compute:

$$Z_{t,i}^\theta = m_t(x_t, \varepsilon_i, \theta), \quad i = 1, \dots, N. \quad (18)$$

These N simulated i.i.d. random l -multiples (triples in case of the NKM), $\{Z_{t,i}^\theta\}_{i=1}^N$, follow the target distribution by construction: $Z_{t,i}^\theta \sim p_t(\cdot|x_t; \theta)$, and thus, we can use them to estimate the conditional density $p_t(z|x; \theta)$ via a standard kernel approximation method. Let us define:

$$\hat{p}_t(z_t|x_t; \theta) = \frac{1}{N} \sum_{i=1}^N K_H(Z_{t,i}^\theta - z_t), \quad (19)$$

where $K_H(\psi) = K(\psi/\sqrt{H})/\sqrt{H}$, $K : \mathbb{R}^l \mapsto \mathbb{R}$ is a generic kernel function that is a symmetric multivariate density and H is a symmetric positive definite bandwidth $l \times l$ matrix. Under regularity conditions on the conditional density p_t and kernel K (Kristensen and Shin, 2012, conditions A.2, A.4, K.1, and K.2, pg. 80–81), we obtain:

$$\hat{p}_t(z_t|x_t; \theta) = p_t(z_t|x_t; \theta) + O_P(1/\sqrt{Nh^l}) + O_P(h^2), \quad N \rightarrow \infty, \quad (20)$$

where the last two terms are $o_P(1)$ if $h \rightarrow 0$ and $Nh^l \rightarrow \infty$.

Using the simulated conditional density $\hat{p}_t(z_t|x_t; \theta)$, we are able to derive the SMLE of θ :

$$\hat{\theta} = \arg \underbrace{\max}_{\theta \in \Theta} \hat{L}_T(\theta), \quad (21)$$

where $\hat{L}_T(\theta) = \sum_{t=1}^T \log \hat{p}_t(z_t|x_t; \theta)$. We use the same set of draws from $\mathcal{F}_\varepsilon(\cdot), \{\varepsilon_i\}_{i=1}^N$ for all values

¹Moreover, the usual assumptions for the consistency and asymptotic normality of the MLE in stationary and ergodic models are imposed on the actual log-likelihood function $L_T(\theta)$ and the associated MLE to ensure that the actual, yet infeasible, MLE $\tilde{\theta}$ is asymptotically well-behaved.

of θ and across t . If $\hat{L}_T(\theta)$ is continuous and differentiable in θ , numerical optimization is facilitated. Considering equation (19), if K and $\theta \mapsto m_t(x_z, \varepsilon_t, \theta)$ are $r \geq 0$ continuously differentiable, the same holds for $\hat{L}_T(\theta)$.

Under the regularity condition, the fact that $\hat{p}_t(z_t|x_t; \theta) \xrightarrow{P} p_t(z_t|x_t; \theta)$ also implies that $\hat{L}_T(\theta) \xrightarrow{P} L_T(\theta)$ as $N \rightarrow \infty$ for a given $T \geq 1$. Thus, the SMLE, $\hat{\theta}$, retains the same properties as the infeasible MLE, $\tilde{\theta}$, as $T, N \rightarrow \infty$ under suitable conditions.

3.1. Important Properties of the SMLE

Kristensen and Shin (2012, pg. 85) argue that the main advantage of the SMLE is its “general applicability”. Starting with observables, the density estimator based on i.i.d. draws is not affected by potential dependence structures in the data, and the SMLE works even if the observations z_t are non-stationary. More important is the potential disadvantage of the SMLE: a curse-of-dimensionality with respect to the dimension of the number of observables, as we smooth only over z_t . Generally, for multi-dimensional models, the estimation performance deteriorates as $l \equiv \dim(z_t)$ increases. We devote careful attention to this issue and extensively study the estimation performance of the SMLE for the three-equation NKM in Section 4. Importantly, however, SMLE does not suffer from the usual curse-of-dimensionality associated with kernel estimators, as largely discussed in Kristensen and Shin (2012). The variance component of the resulting estimator does not need to be controlled by an unbearably larger number of simulations, as the summation in equation (21) reveals an additional smoothing effect and $\hat{L}_T(\theta)$ is recovered at a standard parametric rate of $1/N$. Hence, the curse-of-dimensionality remains only of order $l \equiv \dim(z_t)$, and SMLE will behave similarly to other estimation techniques including MLE in this respect.

On the other hand, given the kernel estimation method and asymptotic properties, the simulated $\hat{L}_T(\theta)$ is generally a biased estimate of the actual $L_T(\theta)$ for fixed approximation precision N and bandwidth $H > 0$. Only $N \rightarrow \infty$ and $H \rightarrow 0$ imply asymptotic consistency. Careful attention thus needs to be devoted to the selection of the bandwidth h with respect to a simulation size and a specific sample of data. Fortunately, in a simulation study, Kristensen and Shin (2012) demonstrate that the SMLE performs well using broad range of bandwidths. A standard identification assumption for the stationary case requires $\mathbb{E}[\log p(y_t|x_t, \theta)] < \mathbb{E}[\log p(y_t|x_t, \theta_0)]$, $\forall \theta \neq \theta_0$. Altissimo and Mele (2009) argue that under its stronger version, the specific choice of bandwidth h is even less important because one can prove the consistency for any fixed $0 < H < \bar{H}$ for some $\bar{H} > 0$ as $N \rightarrow \infty$. This suggests that the proposed methodology is robust to the choice of h from a theoretical perspective because one will assuredly identify well model parameters in large finite samples after $\bar{H} > 0$ is set. However, in a practical application, one still needs to know the threshold level of $\bar{H} > 0$ that can be examined through simulations. In addition to a proper selection of N and H , the kernel K itself needs to belong to a rather broad class of so-called bias reducing kernels. For instance, the Gaussian kernel belongs to this class if the conditional density p has at least two derivatives. Higher numbers of derivatives then facilitate a faster rate of convergence and determine the degree of bias reduction for the estimated conditional density \hat{p} .

With respect to additional theoretical properties, Kristensen and Shin (2012) demonstrate that the SMLE $\hat{\theta}$ is first-order asymptotically equivalent to the infeasible MLE $\tilde{\theta}$ under a set of general conditions satisfied by most models and ensuring that $\hat{p} \rightarrow p$ sufficiently fast, allowing even for mixed discrete and continuous distributions and non-stationarity of the dependent variables. A set

of regularity conditions (A.2, A.4, K.1, and K.2, pg. 80–81) is defined to satisfy these conditions for uniform convergence rates of kernel estimators stated in [Kristensen \(2009\)](#). Moreover, under additional assumptions including, e.g., stationarity, results regarding the higher-order asymptotic properties together with expressions for the bias and variance components of the SMLE $\hat{\theta}$ due to kernel estimation and numerical simulations compared to the actual MLE are derived.

4. Monte Carlo Simulations

Prior to empirical estimation of the NKM parameters,² we pursue a rigorous numerical investigation of the small sample estimation performance of the SMLE method for the NKM and of issues such as sources of bias and estimation uncertainty potentially rooted in the curse of dimensionality, various specifications of the model, or a choice of approximation precision N and bandwidth $h > 0$. A computationally extensive simulation study allows us to determine the extent to which the estimation strategy is able to recover the pseudo-true parameters in the controlled environment using the pseudo-true data generated from the model.

4.1. Simulation Setup

This section describes a general setup for all the subsequent numerical exercises if not explicitly stated otherwise. All computations are conducted using MATLAB version R2015b (8.6.0.267246). The computational burden of the estimation procedure has been made manageable utilizing MATLAB [Parallel Computing Toolbox](#) in combination with the `parfor` command that executes for-loop iterations in parallel. Starting with the NKM model, we first calibrate variables that have strong support in the related NKM empirical literature. The discount factor is set to $\nu = 0.99$ based on a large number of studies in which the same value comes from empirical estimation. No memory of the past forecast performance is assumed for the BR model ($\rho = 0$) based on the results of [Jang and Sacht \(2016, 2018\)](#), who generally find this parameter insignificant.³ Parameters for the numerical study are calibrated according to recent results obtained by [Jang and Sacht \(2017, Table 3, PEB scenario\)](#); however, all qualitative results presented are robust to the calibration employed. Parameter pseudo-true values are summarized in [Table 1](#).

We study the estimation performance of the SMLE at various updating frequencies represented by the parameter of the intensity of choice set to $\gamma = \{0.1, 1, 10\}$. The suggested range of values resembles an economic intuition of increasing intensities of switching of agents between forecasting strategies ranging from almost negligible ($\gamma = .1$) through moderate ($\gamma = 1$) to strong ($\gamma = 10$).⁴

²The three-equation NKM represents a realistic macroeconomic agent-based model with multivariate observables. This allows us to study the capability of the SMLE in a more complex framework compared to univariate applications to estimate financial agent-based models developed by [Kristensen and Shin \(2012\)](#), who estimate the short-term interest rate model of [Cox et al. \(1985\)](#), the jump-diffusion model of daily stock returns of [Andersen et al. \(2002\)](#), and generic Markov decision processes, and [Kukacka and Barunik \(2017\)](#), who estimate the [Brock and Hommes \(1998\)](#) heterogeneous agent model.

³An additional robustness check regarding the memory parameter in the forecast performance, ρ , is provided in [Subsection 4.2.4](#).

⁴Our selection of numerical values of γ also closely follows the related literature. For comparison, [De Grauwe \(2010, 2011, 2012\)](#) and [Liu and Minford \(2014\)](#) calibrate $\gamma = 1$, [Jang and Sacht \(2016, 2018\)](#) employ a range $\gamma = \{0.1, 1, 10, 100\}$ for the [De Grauwe \(2011\)](#) model, while [Grazzini et al. \(2017\)](#) use $\gamma = 5$ when estimating the [De Grauwe \(2012\)](#) model. We note that a negative γ lacks any economic sense, as it would imply illogical switching towards less precise forecast strategies. On the other hand, high values of γ cause an unrealistic speed of belief updating that is unlikely in reality. Indeed, [Jang and Sacht \(2016\)](#) use $\gamma = 100$ as an approximation to infinity.

The results thus provide a sensitivity analysis w.r.t. the intensity of choice parameter γ . The idiosyncratic shocks ε_s are drawn from a normal distribution with mean zero and variance σ_s^2 , $s = \{y, \pi, r\}$, that is a standard and reasonably realistic assumption for \mathcal{F}_ε in equation (15). We can thus exploit favorable theoretical properties of the Gaussian kernel (Kristensen and Shin, 2012, pg. 81) related to equation (19).

In Monte Carlo simulation, we study the estimation performance of the SMLE under three lengths of resulting time series generated from the model: 120, 220, and 420. This allows us to obtain important knowledge about asymptotic tendencies of the estimator with increasing sample size while maintaining an admissible time span for common quarterly macroeconomic data. Moreover, 100 additional initial simulated observations are always discarded as a burn-in period in which the dynamic properties of the model are established. The time-dependent unconditional volatility in the output gap $\lambda_{y,t}$ is computed as a moving standard deviation over last 20 simulated periods, representing 5 years of a stylized business cycle, with a reference standard deviation computed from the empirical dataset for the first 20 transient periods. The random seed is controlled in each simulation and set as MATLAB `rng('default')`.

With regard to the setup of the multivariate SMLE method, we apply a constrained joint multivariable function estimation of 10 and 7 parameters for the BR and the RE model, respectively. The parameter space is restricted to intervals summarized in Table 1. These constraints are based on either theoretical borders for given parameters (cf. Section 2) or empirical estimates from recent literature (Jang and Sacht, 2016, 2018) combined with the results of a preliminary rough search based on a broader space. The estimation method is, nonetheless, always ensured to be given sufficient leeway, especially for the BR parameters (β, δ, γ) . The first 20 transient periods are not considered for the optimization procedure, thereby decreasing the effective length of the simulated series to $T = \{100, 200, 400\}$. This makes manipulation with simulated data identical to that with empirical datasets. The precision of the kernel density estimation is set to $N = 1000$. Moreover, to tackle potential numerical limitations of the optimization procedure, the number of initial conditions randomly generated from given intervals is set to 5, and only the result with the greatest simulated log-likelihood is considered. To ensure the statistical validity of the results while keeping the computational burden manageable, 300 independent runs are always conducted. We carefully tested for sufficiency of the suggested setting for N , the number of initial conditions, and the number of runs. For the estimation of the conditional density $p_t(z|x;\theta)$, we consider the multivariate Gaussian kernel and Silverman’s (1986) rule of thumb to set an optimal bandwidth matrix: $\sqrt{H_{s,s}} = (4/[(l+2)N])^{1/(l+4)} \hat{\sigma}_s$, where $l = 3$, $s = \{y, \pi, r\}$, and $\hat{\sigma}_s$ denotes the standard deviation of the of the s^{th} variable of $\{Z_{t,i}^\theta\}_{i=1}^N$, and off-diagonal terms $H_{s_1,s_2} = 0$, $s_1 \neq s_2$.

4.2. Numerical Results

In the numerical study, we investigate the extent to which the SMLE method is able to recover the pseudo-true parameter values of the NKM in small samples. We are interested primarily in tendencies to consistency and asymptotic efficiency of the estimator as well as potential bias and its sources. A comparison between the estimation performance of the BR model and the RE model is provided. We present overall results in Table 1 and a graphical depiction for the BR model with the moderate value $\gamma = 1$ and for the RE model in Figure 1. In the Appendix, Figure 5 reports additional graphical results for other numerical values $\gamma = \{0.1, 10\}$ to provide a complete overview.

In the following text, we also focus on rather subtle differences in estimation performance for these different values of the intensity of choice $\gamma = \{0.1, 1, 10\}$.

4.2.1. Estimation Performance

The results of the simulation study in [Table 1](#) reveal promising estimation performance of the SMLE in the multivariate setting for the BR model. Similar information as in [Table 1](#) for $\gamma = 1$ is then depicted visually in [Figure 1](#) (panels a, c, and e). First, the structural parameters of the NKM— τ , κ , ϕ_π , and ϕ_y —are generally estimated very precisely. The Taylor rule coefficient ϕ_π values are recovered nearly perfectly. Moreover, the slope of NKPC κ reveals a very good ability to be accurately estimated through all studied settings. The only small exception is observed for the elasticity of substitution in consumption τ , measuring the inherited persistence in terms of cross-volatility within the inflation-output gap nexus; although we do obtain unbiased point estimates, the 95% confidence intervals are very wide for shorter time series with $T = \{100, 200\}$, especially for small $\gamma = 0.1$ ([Figure 5](#), panels a and c). However, this is a natural behavior of the estimator since SMLE generally requires a sufficient sample to work properly ([Kukacka and Barunik, 2017](#)), and a small intensity of choice $\gamma = 0.1$ represents a case in which the adaptive switching between strategies is considerably restrained, which might hinder model dynamics and the variability crucial for estimation. This seems to be a crucial piece of knowledge gained for the interpretation of the empirical results. More important, larger samples and richer dynamics of the model under higher γ strongly supports the estimation performance of τ (see [Figure 5](#), panel f) making it comparable with the results for κ and thus suggesting favorable asymptotic tendencies of the estimator.

The BR parameters— β , δ , and γ —are generally more challenging to estimate. The subjective mean of the future output gap β exhibits good estimation performance but, like τ , presents large 95% confidence intervals for the shortest time series with $T = 100$ and for small $\gamma = 0.1$ (e.g., [Figure 5](#), panel a). Again, in all other cases, β is estimated relatively precisely with no apparent bias. On the other hand, the parameter measuring the degree of divergence δ suffers from a general tendency towards statistical insignificance in all simulations under the SMLE. This estimation issue is due to the interaction of δ with the time-varying unconditional standard deviation $\lambda_{y,t}$, see equations (5) and (6), that naturally hinders the estimation precision of the SMLE for this parameter. We further examine this issue in [Subsection 4.2.4](#), where we estimate a simplified model with a fixed $\lambda_{y,t} = \text{const}$. We observe markedly increased precision of δ estimates in that case; however, these are accompanied by newly emerged large biases for both β (−) and δ (+).

Table 1: Results of the Monte Carlo simulations

Param.	Calib.	BR NKM									RE NKM		
		$\gamma = .1$			$\gamma = 1$			$\gamma = 10$			100	200	400
		100	200	400	100	200	400	100	200	400			
τ	.32	.57	.44	.37	.36	.38	.35	.31	.34	.32	.61	.62	.52
(0, 1)		(0-1)	(.02-1)	(0-1)	(.02-.99)	(.11-.99)	(.18-.58)	(.11-.68)	(.22-.56)	(.21-.44)	(0-1)	(.11-1)	(0-1)
κ	.23	.35	.33	.28	.22	.23	.23	.23	.23	.22	.40	.47	.39
(0, 1)		(0-1)	(0-1)	(0-.95)	(.06-.37)	(.14-.34)	(.17-.32)	(.13-.35)	(.15-.32)	(.17-.28)	(0-1)	(0-1)	(0-1)
ϕ_π	1.57	1.64	1.66	1.63	1.57	1.58	1.59	1.56	1.57	1.57	1.72	1.79	1.75
(0, 3)		(1.26-2.13)	(1.40-2.07)	(1.46-1.86)	(1.32-1.78)	(1.48-1.73)	(1.52-1.66)	(1.51-1.62)	(1.55-1.60)	(1.55-1.60)	(.89-2.22)	(1.43-2.09)	(1.49-1.94)
ϕ_y	.34	.30	.28	.26	.32	.32	.31	.34	.33	.33	.25	.26	.23
(0, 1)		(0-.80)	(.06-.68)	(.03-.45)	(.12-.48)	(.19-.40)	(.26-.42)	(.27-.44)	(.26-.42)	(.28-.37)	(0-.72)	(.02-.69)	(.03-.45)
σ_y	.47	.30	.29	.30	.25	.27	.29	.24	.27	.29	.33	.34	.33
(0, 1)		(17-.55)	(.19-.49)	(.23-.50)	(.14-.44)	(.17-.39)	(.22-.38)	(.10-.39)	(.17-.35)	(.22-.36)	(.20-.60)	(.20-.52)	(.23-.51)
σ_π	.33	.20	.21	.20	.18	.20	.20	.19	.20	.20	.21	.23	.22
(0, 1)		(.07-.40)	(.15-.34)	(.16-.30)	(.08-.28)	(.14-.25)	(.16-.24)	(.12-.26)	(.14-.24)	(.15-.24)	(.13-.35)	(.17-.35)	(.17-.32)
σ_r	.29	.06	.08	.09	.10	.10	.10	.08	.10	.11	.05	.06	.07
(0, 1)		(0-.16)	(0-.18)	(0-.16)	(0-.15)	(0-.17)	(0-.17)	(0-.18)	(0-.17)	(0-.16)	(0-.15)	(0-.17)	(0-.15)
β	2.16	.55	.87	1	2.05	2.14	2.23	2.21	2.42	2.21	-	-	-
(0, 4)		(0-4)	(0-4)	(0-4)	(1.01-4)	(1.34-3.15)	(1.67-2.87)	(.83-3.47)	(1.65-3.37)	(1.61-2.72)	-	-	-
δ	.59	.59	.69	.39	.94	.76	.56	.61	.41	.54	-	-	-
(0, 3)		(0-3)	(0-3)	(0-3)	(0-2.89)	(0-1.67)	(.05-1.33)	(0-1.87)	(0-1.08)	(0-1.28)	-	-	-
γ		.37	.42	.26	1.17	1.04	.99	13.03	10.35	11.49	-	-	-
(0, 20)		(0-11.49)	(0-5.23)	(.02-2.64)	(.47-13.20)	(.60-4.42)	(.69-1.76)	(2.79-20)	(3.12-20)	(3.75-20)	-	-	-

Constrains for optimization and for initial conditions given in $\langle \rangle$ brackets. T denotes the length of the executed time series. Sample medians reported, and 95% confidence intervals of sample estimates reported in $()$ brackets. Figures rounded to 2 decimal places.

The most important result is the estimation performance of the switching coefficient γ . Even in simple univariate financial models, from which the concept of the behavioral switching governed by the intensity of choice via the multinomial logit model is derived, capturing the effect of this coefficient is generally difficult (Boswijk et al., 2007; Hommes, 2013). In recent estimation studies regarding the BR NKM, Liu and Minford (2014) estimate $\hat{\gamma} = 0.85$ for the De Grauwe (2010) BR model specification using the Indirect Inference method with VAR as the auxiliary model but strongly reject the BR model overall in favor of rational expectations for US data ranging from 1981:Q4 to 2013:Q4. Their results also suggest that “the problem lies with the expectation scheme itself and not with its parameter values” (pg. 414). Jang and Sacht (2016, 2018) leave γ unestimated for the De Grauwe (2011) model and follow a grid strategy in which they compare results when varying the intensity of choice between 0.1 and 100. Grazzini et al. (2017) employ a Bayesian method to estimate the De Grauwe (2012) version of the BR NKM and show that the posterior distribution coincides with the prior distribution, which suggests that the data and the estimation procedure do not generate any additional knowledge about the distribution of the γ parameter. On the contrary, our results show that the SMLE is well capable of estimating the pseudo-true intensity of choice γ under the discrete choice expectation updating scheme (9). In other words, compared to other estimation methods applied to date and assuming the BR NKM, in this simulation study, the SMLE demonstrates a good ability to precisely estimate γ , i.e., to detect signs of behavioral switching in the model-generated output data. Specifically, we observe excellent estimation performance for a moderate ($\gamma = 1$) intensity of choice; see Figure 1 (panels a, c, and e). Additional robustness checks with a weaker ($\gamma = 0.5$) and a stronger ($\gamma = 2$) pseudo-true intensity of choice that strongly support this conclusion are not reported for the sake of space but are available from the authors upon request. As we found above for κ and τ , insufficient sample size ($T = 100$, Figure 1, panel a) hinders accurate estimation and increases the standard error of the estimate, but with an increased sample size (Figure 1, panels c and e) the results gain considerable precision. For stronger behavioral switching under $\gamma = 10$ (Figure 5, panels b, d, and f), the intense dynamics of the model weaken its ability to accurately estimate the pseudo-true value of γ , but the results still do not suffer from potential statistical insignificance. Based on a considerable number of estimates close to the upper border of the parameter space in this case, one can expect a potential upward bias of the estimator for the parameter γ if wider constraints are introduced.

On the other hand, the estimator is likely to suffer considerably from the curse of dimensionality for the third set of studied parameters, as the simulations reveal a strong downward bias of the magnitude of idiosyncratic shocks represented by their standard deviations σ_y , σ_π , and σ_r . Estimation of σ_r seems to suffer the most, as this parameter is statistically insignificant at the 5% level in all reported cases under standard distributional assumptions. Interestingly, this bias does not seem to be influenced by the actual value of the pseudo-true intensity of choice γ , as it appears stable across all simulations and exhibits only a very subtle tendency to decrease with increasing sample size. We thus suggest that this behavior is not primarily caused by a small sample size or constrained dynamics of the model when the pseudo-true γ is relatively low. To shed additional light on this issue, we test a set of working hypotheses (see a detailed description in Subsection 4.2.4) assigning the origin of this bias to various potential sources such as additional uncertainty brought to the system based on actual switching, an additional non-modeled ‘shock’ introduced through the output gap expectations heuristics in the behavioral model, or the inherited persistence in the case of the rational expectations model. Under all hypotheses, however, the underestimation of shocks

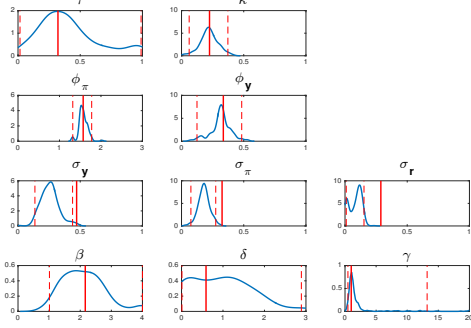
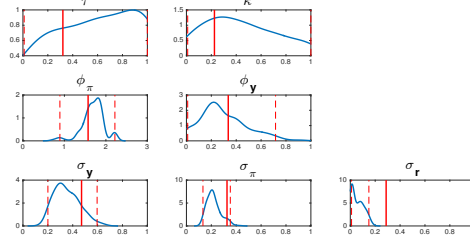
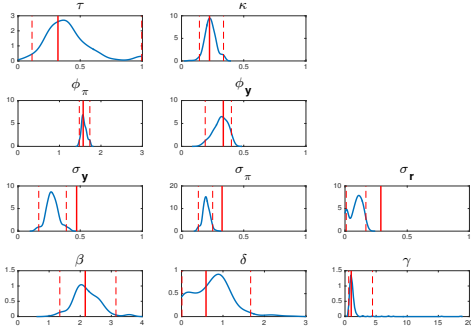
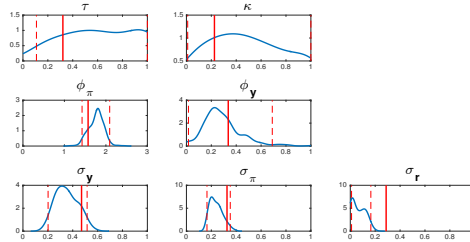
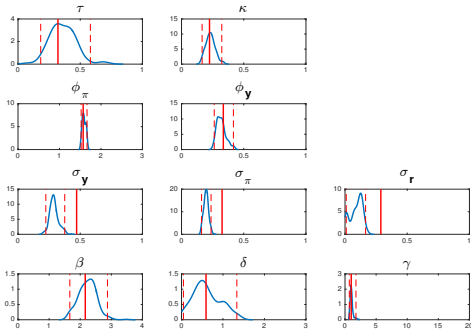
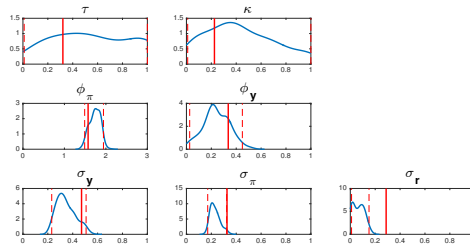
(a) BR model, $\gamma = 1$, $T = 100$ (b) RE model, $T = 100$ (c) BR model, $\gamma = 1$, $T = 200$ (d) RE model, $T = 200$ (e) BR model, $\gamma = 1$, $T = 400$ (f) RE model, $T = 400$ 

Figure 1: Densities of the pseudo-true parameter estimates. Blue curves depict kernel density estimates of the sample densities, bold red lines show the pseudo-true values, and dashed red lines depict 95% intervals of sample estimates.

remains qualitatively unchanged, and we thus conclude that it most likely represents a consistent estimation issue due to the emergent nonlinear complexity of the model that the SMLE method is not able to capture completely. We further stress that one needs to take this information seriously into consideration when interpreting such empirical results.

4.2.2. Asymptotic Tendencies

Tendencies to favorable asymptotic properties of the estimator are apparent with an increasing number of observations for all parameters. We are well aware that a sample size of $T = 400$ can hardly approximate infinity, but we are interested primarily in practical estimation performance for quarterly macroeconomic data, for which a century’s worth of data is already generally unavailable.⁵

We generally observe a considerable narrowing of the 95% confidence intervals of the sample estimates when T goes from 100 to 200 and further to 400. This tendency appears weak for the case of $\gamma = 0.1$ (Figure 5, panels b, d, and f), especially for τ , κ , standard deviations of shocks (σ_y , σ_π , σ_r), β , and δ . This suggests that the constrained model dynamics also influence the overall asymptotic properties of the estimator. The results demonstrate a surprisingly large effect of an increasing sample size generally for the estimation precision of the intensity of choice γ but only for its moderate value ($\gamma = 1$; see Figure 1, panels a, c, and e).⁶ On the other hand, almost no effect is observed for its high value ($\gamma = 10$, Figure 5, panels b, d, and f). From a theoretical perspective, the very subtle tendency towards bias reduction for standard deviations of shocks with increasing sample size is favorable, but the speed with which the estimates approach the true value is slow, which is of practical importance for empirical macroeconomic time series estimation.

4.2.3. Shape of the Log-Likelihood Function

A set of regularity conditions A.1-A.4 imposing restrictions on the model and the conditional density is defined in Kristensen and Shin (2012) such that the estimated conditional density \hat{p} converges sufficiently fast to the true conditional density p , and hence, the asymptotic equivalence of the estimated $\hat{\theta}$ and the true θ parameter vectors is assured. Those authors assert that these assumptions are “quite weak and are satisfied by many models” (Kristensen and Shin, 2012, pg. 81); nonetheless, the analytical intractability of the analyzed behavioral macroeconomic model does not allow us to mathematically verify these conditions. Therefore, we exploit the computational approach and verify the smoothness condition, identification of parameters, and existence of a unique optimum by assessing the simulated log-likelihood functions via graphical tools.

Figure 2 shows the shape of the simulated log-likelihood functions for the BR model with $\gamma = 1$ (panels a and c). As we are unable to graphically depict a 11-dimensional object, the individual curves show transversal profiles of the simulated likelihood function in planes of given parameters. Other parameters are held fixed at their estimated values from Table 1. A smooth shape of the surface resulting from the use of a Gaussian kernel based on a sufficient number $N = 1000$ of approximation points is clearly observable for all dimensions over the entire domain. Unique

⁵In addition, the estimation procedure’s extensive computational burden does not allow us to run a representative Monte Carlo experiment that maintains the suggested setting with sample sizes of higher orders of magnitude within a reasonable time even using high-speed servers for parallel computing.

⁶We again verified robustness of this finding via checks with $\gamma = \{0.5, 2\}$. The results are available from the authors upon request.

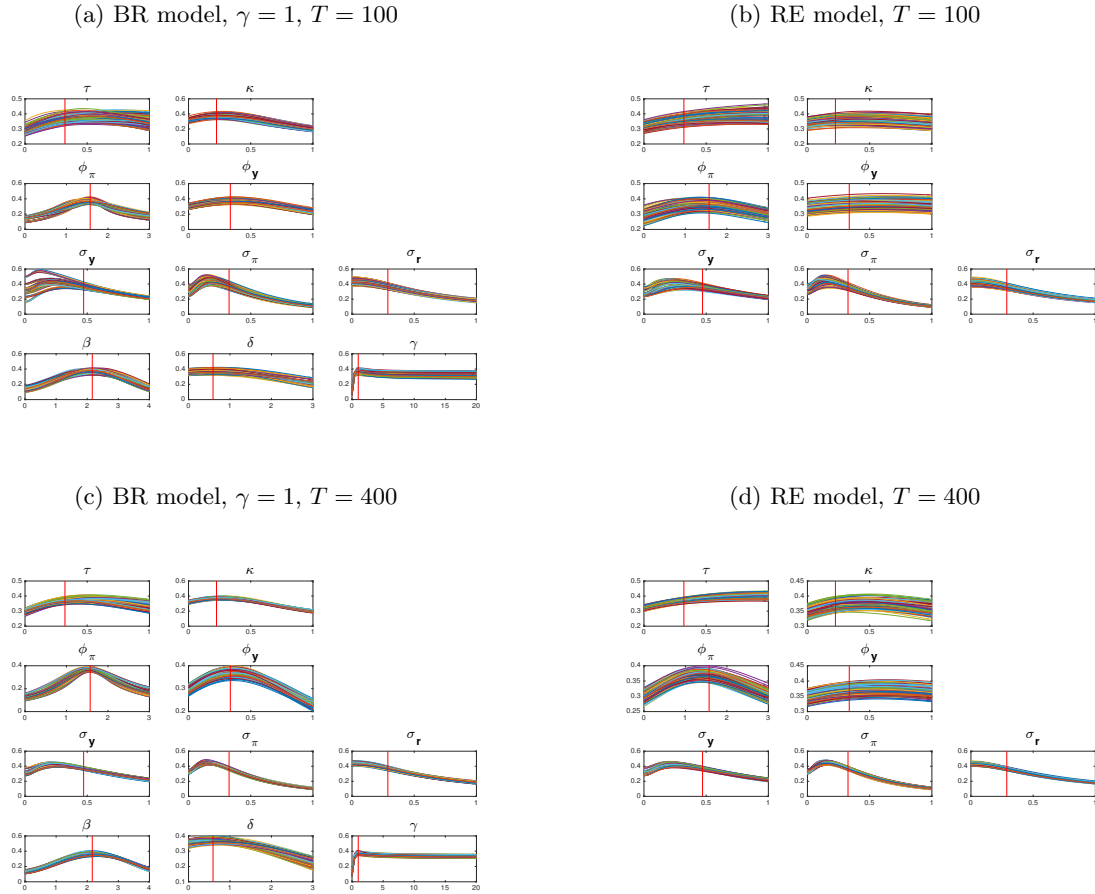


Figure 2: Profiles of the simulated log-likelihood function. Results based on 100 runs. Bold red lines show the pseudo-true values.

maxima are clearly detectable for the majority of parameters with the most spiky shape in the direction of parameter ϕ_π , naturally reflecting the numerical results reported in [Table 1](#) and [Figure 1](#). Conversely, a very flat surface of the likelihood function for a large part of the domain can be observed for the behavioral coefficients δ and γ , confirming our results regarding the challenging identification of these parameters. The bias of σ_y , σ_π , and σ_r is also obvious from the shape of the log-likelihood. When increasing the sample size from $T = 100$ (panel a) to $T = 400$ (panel b), we surprisingly do not obtain distinctively sharper shapes but rather consistent behavior of the log-likelihood for all 100 independent runs. Differences between independent runs are observable primarily for parameters τ and σ_y . Especially for τ , such a feature explains the poor estimation performance for short time series with $T = 100$ ([Figure 1](#), panel a). For σ_y , interestingly, we observe considerable dispersion of the shapes around the maximum for small sample size $T = 100$ ([Figure 2](#), panel a), suggesting its sensitivity to different random seeds. We also verified that all reported qualitative results are robust to the particular value of the intensity of choice γ imposed. In summary, based on smooth surfaces and unique maxima observed in all directions of the simulated log-likelihood function, we can generally assume that the regularity conditions and identification

of parameters are satisfied for the BR NKM.

4.2.4. Comparison with the RE Model and Other Specifications

When we compare the BR model’s estimation performance with that of the RE model, the SMLE method interestingly provides more accurate estimation of the behavioral specification. This result is surprising, since the RE specification is based on a simpler dynamic system than the BR. However, the simulated data from the RE NKM are characterized by large fluctuations exhibiting a low level of persistence. In this particular case, the SMLE exhibits better performance for a more structural data set that contains more information on the system. This could fit a particular case of the BR specification with more inherited persistence in model dynamics.

Simulation-based results for the RE model are summarized in the right column of [Table 1](#) and in [Figure 1](#) (panels b, d, and f). Although only 7 parameters are estimated simultaneously for the RE model, the estimation performance is generally worse than that of the more complicated BR model due to qualitatively different dynamics of the output time series of the RE models compared to BR models. While a BR model output displays a considerable stability and path dependency emerging from incorporation of behavioral heuristics for forecasting output and inflation gaps, fluctuations of the output time series of a RE model are, on the other hand, more intense because the notion of path dependence is missing from the baseline, purely forward-looking specification. Exogenous random shocks rather than (cross-)autocorrelation structures are then the main driver of the RE NKM model dynamics, which naturally stymies proper detection of pseudo-true parameter values for any estimation method. The inability of the SMLE to approximate the likelihood function accurately enough for strongly fluctuating time series is widely studied in [Kukacka and Barunik \(2017\)](#). This is because the SMLE criterion functions then display a flatter surface, which hinders the optimization algorithm from reaching the global optimum in a multidimensional parameter space. Moreover, its shape is likely to be more strongly influenced by the initial random conditions, thereby increasing the uncertainty of estimates from a Monte Carlo study. This can be generally observed in [Figure 2](#) (panels b, and d). The individual courses of depicted shapes are more dispersed, and the maxima of the likelihood functions are more complicated to detect; see, e.g., the likelihood profiles for τ that in many cases seem to be increasing for the whole domain, for all structural parameters of the RE model— τ , κ , ϕ_π , and ϕ_y —compared to panels (a) and (c) depicting similar results for the BR model. Another related issue is captured by the shapes of the likelihood functions for ϕ_π and ϕ_y that alone visually suggest the opposite biases to those documented by the joint estimation in [Figure 1](#) (panels b, d, and f). This additional seemingly puzzling inconsistency simply suggests a more complex problem of identification when optimizing the joint log-likelihood function for the RE model.

Specifically, we observe an overall bias not only for standard deviations of shocks but also for all structural parameters. Moreover, the bias does not tend to disappear as the sample size increases, suggesting unfavorable asymptotic tendencies of the estimator in the case of the RE model. In terms of efficiency, the 95% sample estimate intervals are very large for τ and κ , large for the Taylor rule coefficients ϕ_π and ϕ_y , and comparable for σ_y , σ_π , and σ_r . The relative improvement in efficiency as sample size increases is generally slower for all parameters except ϕ_π than in the BR model.

We also attempt to estimate ρ from equation (8) as the 11th parameter. [Jang and Sacht \(2016\)](#)

found this agent memory coefficient for computing forecast performance measure to be statistically insignificant from estimation via the SMM, to which we contrast the SMLE. [Grazzini et al. \(2017\)](#) also reported it to have a negligible effect for Bayesian estimation, as the posterior distribution did not depart significantly from the prior distribution. Our estimation yields comparable results (not reported) to those of [Grazzini et al. \(2017\)](#), suggesting that either the model behavior or the estimation method is insensitive to the actual pseudo-true value of the memory parameter because resulting point estimates generally appear to be concentrated near the middle point 0.5 of the parameter space interval $\langle 0, 1 \rangle$ for a variety of considered pseudo-true $\rho = \{0, 0.2, 0.8\}$.

Finally, we test a set of working hypotheses assigning the consistent estimation bias for standard deviations of shocks σ_y , σ_π , and σ_r to three potential sources: (i) possible additional uncertainty brought to the system by the expectation updating scheme (9), (ii) an additional non-modeled ‘shock’ introduced to the system through the output gap expectations heuristics, specifically via the time-varying unconditional standard deviation $\lambda_{y,t}$ interacting with the degree of divergence δ (see equations (5) and (6), in the behavioral model), or (iii) the inherited persistence in case of the rational expectations model. All potential sources of additional uncertainty might reduce the importance of the exogenous random shocks for the overall dynamics of the model and the variability of model output. We thus simulate the model with (i) a fixed $\gamma = 0$, (ii_a) a fixed $\lambda_{y,t} = \text{const}$, (ii_b) fixed β and δ for the BR model, or (iii) fixed τ and κ for the RE model. Under these hypotheses, we simplify the model somewhat, expecting a reduction in the bias for the shocks, while also being interested in additional side effects in terms of improved/reduced estimation performance of the SMLE for the model. However, under all three hypotheses (results not reported), the underestimation of shocks remains qualitatively unaffected, while a positive bias for δ and an overall deterioration of the estimation precision for all structural parameters emerge in (i), biases for both β (−) and δ (+) emerge in (ii_a), a negative bias for τ appears in (ii_b), and an increased bias for shocks emerges in (iii). This initially rather surprising reduced estimation performance of the simplified models can be explained similarly as that of the RE model at the beginning of this section. All our hypotheses are thus rejected, and we conclude that a nonlinear complexity of the model that the SMLE method is not able to capture completely is most likely the cause of the reported bias for the shocks.

5. Empirical Applications

This section examines the empirical performance of the SMLE in a multivariate setting, which considers both the purely forward-looking and hybrid versions of the model. The inclusion of the latter relies on the empirical fact that the RE NKM cannot capture the inertia in inflation and output since it does not include backward-looking components if no autocorrelation in the shocks is assumed. In the macroeconomic literature, this is termed the so-called ‘(inflation) persistence problem’ by [Chari et al. \(2002\)](#), among others. Since the selected data for the Euro Area and the US economy exhibit a high degree of persistence, we investigate the empirical performance of the NKM in its hybrid version. In addition, we connect with our results obtained in [Subsection 4.2.2](#) and shed light on the estimation outcomes for the forward-looking version given multiple sets (for the US only) that differ in the number of observations included. In the subsequent empirical section, we therefore estimate the RE and BR NKM using Euro Area data from 1975:Q1 to 2009:Q4 with 140 observations. For the US economy, we broadly consider three different periods, i.e., the *Great*

Inflation (GI, 1965:Q1–1979:Q4, 60 observations), the *Great Moderation* (GM, 1980:Q1–2009:Q4, 120 observations), and the *Great Moderation+Great Recession* (GM+GR, 1980:Q1–2016:Q4, 148 observations).

5.1. Estimation Setup

The empirical estimation algorithm generally adopts the setup of the simulation study but with two minor alterations that are described below. First, we report results based on 500 runs. Second, compared to the parameter space for the numerical study, we restrict the interval for the intensity of choice γ by 10 from above based on a preliminary rough search. We further bound the interval for the ϕ_π by 1 from below following a standard calibration from the literature to support optimization. We set, again, $\nu = 0.99$ and $\rho = 0$ based on the empirical evidence reported in a overwhelming number of studies and according to our discussion in [Subsection 4.2.4](#), respectively. The range of parameter values is completely described in [Table 2](#) and [Table 3](#).

5.2. Data

The behavioral macroeconomic models are estimated using Euro Area and US data sets. We retrieve the Euro Area data set from the 10th update of the Area-Wide Model quarterly database (see [Fagan et al., 2001](#)).⁷ The sample spans the period from 1975:Q1 to 2009:Q4 with 140 observations. The GDP deflator is used to measure inflation in the Euro Area. The short-term nominal interest rate and real GDP are used to measure the gaps in the nominal interest rate and output in the Euro Area.⁸ A standard smoothing parameter of $\lambda = 1600$ is used to estimate the trend of the observed data from the Hodrick-Prescott filter for output, inflation, and the nominal interest rate.

We take the US data from the website of the Federal Reserve Bank of St. Louis.⁹ Inflation is measured using the seasonally adjusted GDP deflator with 2009 as the base year. Output is obtained from seasonally adjusted real GDP based on billions of chained 2009 dollars. The effective federal funds rate is used to measure the short-term nominal interest rate in the US. The sample covers the period 1965:Q1 to 2016:Q4. As mentioned above, the data set includes three different subperiods, i.e., GI [US(60)], GM [US(120)], and GM+GR [US(148)]. The different subperiods cannot be considered for the Euro Area due to limited data availability and the lack of structural breaks in inflation volatility.

5.3. Empirical Results: Forward-Looking RE vs. BR

The parameter estimates for the forward-looking case are shown in [Table 2](#). [Figure 3](#) (panels a-d) depicts the associated densities of parameter estimates. First, it is immediately apparent that in the BR and RE NKM, a number of parameters are estimated to be insignificant. Based on the Euro Area data this observation is accompanied by point estimates of zero for σ_y (BR) and σ_r (RE) as well as κ (both specifications). For the US data, we find insignificant estimates of the majority of parameters for the BR NKM. This holds for τ , κ , σ_y , σ_r and δ across all subperiods, ϕ_y only for both US(120) and US(148), σ_π only for US(120) and β only for US(148). In the RE NKM

⁷Available at: www.eabcn.org/page/area-wide-model [Accessed 10 August 2018].

⁸The time series in the Area-Wide Model database have the following abbreviations: GDP deflator = ‘YED’, short-term nominal interest rate = ‘STN’ and real GDP = ‘YER’.

⁹Available at: fred.stlouisfed.org [Accessed 10 August 2018].

Table 2: Results of the empirical study: forward-looking RE vs. BR

Par.	BR NKM				RE NKM			
	Euro (140)	US (60)	US (120)	US (148)	Euro (140)	US (60)	US (120)	US (148)
τ	.68	.52	.86	.05	1	1	1	1
$\langle 0, 1 \rangle$	(.66-.71)	(0-1)	(0-1)	(0-1)	(.76-1)	(1-1)	(.56-1)	(1-1)
κ	0	.64	.27	.23	0	.25	.50	.50
$\langle 0, 1 \rangle$	(0-0)	(0-.76)	(0-1)	(0-1)	(0-0)	(.24-.26)	(.48-.51)	(.49-.51)
ϕ_π	2.14	2.06	1	1.34	1	1	1	1
$\langle 1, 3 \rangle$	(1.99-2.26)	(1-2.75)	(1-3)	(1-3)	(1-1)	(1-1)	(1-2.99)	(1-1.02)
ϕ_y	.78	.19	.70	.54	.72	0	0	0
$\langle 0, 1 \rangle$	(.77-.81)	(.10-1)	(0-1)	(0-1)	(.70-.76)	(0-0)	(0-1)	(0-.18)
σ_y	0	0	.03	.02	.96	.23	.30	.30
$\langle 0, 1 \rangle$	(0-.02)	(0-0)	(0-.20)	(0-.23)	(.79-1)	(.17-.59)	(.26-.34)	(.27-.35)
σ_π	.10	.03	.04	.11	.20	0	0	0
$\langle 0, 1 \rangle$	(.09-.12)	(0-.08)	(0-.16)	(.02-.36)	(.14-.25)	(0-0)	(0-0)	(0-0)
σ_r	.08	0	.14	.24	0	.14	.53	.36
$\langle 0, 1 \rangle$	(.06-.11)	(0-.29)	(0-.67)	(0-.67)	(0-.4)	(.10-.19)	(.41-1)	(.28-.50)
β	1.42	.69	3.73	1.88	-	-	-	-
$\langle 0, 4 \rangle$	(1.41-1.45)	(0-3.70)	(.11-4)	(0-4)	-	-	-	-
δ	1.24	2.05	1.68	.41	-	-	-	-
$\langle 0, 3 \rangle$	(1.19-1.27)	(0-2.79)	(0-3)	(0-2.60)	-	-	-	-
γ	7.01	.95	.53	.68	-	-	-	-
$\langle 0, 10 \rangle$	(5.24-9.20)	(.14-9.99)	(0.15-3.29)	(.07-4.39)	-	-	-	-
LL	.17	.30	.17	.14	.07	.05	.07	.09
	(.17-.18)	(.19-.38)	(.07-.08)	(.11-.16)	(.24)	(.04-.05)	(.06-.07)	(.09-.09)

Constraints for optimization and for initial conditions given in $\langle \rangle$ brackets. Sample medians reported, and 95% confidence intervals of sample estimates reported in $()$ parentheses. LL denotes the average log-likelihood of the estimated models and represents statistical fit. Figures rounded to 2 decimal places.

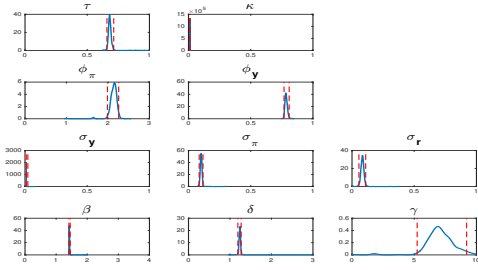
case for the US, only ϕ_y and σ_π are insignificant across all subperiods, where for both shocks, the point estimates are indeed strictly zero.

The troublesome results from the BR NKM are at odds with our analysis in [Subsection 4.2.2](#), where we report the asymptotic tendencies of the pseudo-true parameter estimator as the number of observations is increased. [Figure 3](#) shows bimodal-shaped densities for a large number of parameters. It can be concluded from this that the quantity and quality of the data at hand for all subperiods produce a flat likelihood in each case. Therefore, the interpretation of the results for different monetary policy regimes in the US has some flaws.

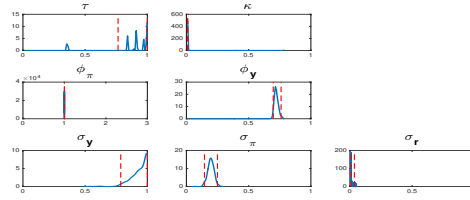
Our observations indicate that because the whole sample period for the US can be divided into the three different subperiods, the BR specification struggles when dealing with the structural break in inflation volatility at the beginning of the Greenspan era of monetary policy. As the suggestion is to consider a long time series of macroeconomic data from, say, 1965:Q1 to 2016:Q4, which would favor the asymptotic tendencies reflected in the parameter estimates given a large number of observations, this can also be crucial. In this regard, [Jang and Sacht \(2016, pg. 103\)](#) already report that the BR model seems to have difficulties in explaining the high volatility in the interest rate and inflation gap over the whole sample period when the SMM approach is applied.

As mentioned in our numerical study, certain parameters are up- or downward biased. The former is true for τ for the US and Euro Area data but with respect to the RE NKM only. A downward bias is observed for the majority of both policy coefficients and all shocks for different data sets

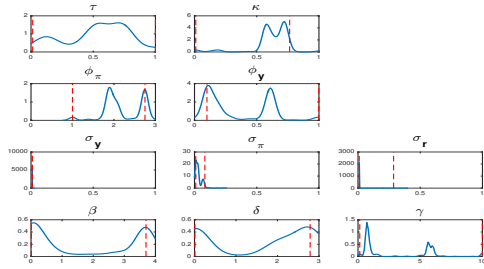
(a) BR model, Euro Area data (140)



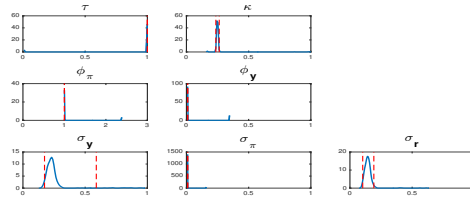
(b) RE model, Euro Area data (140)



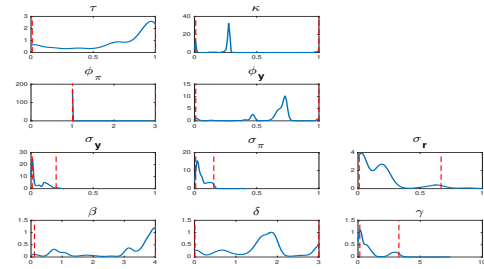
(c) BR model, US data (80)



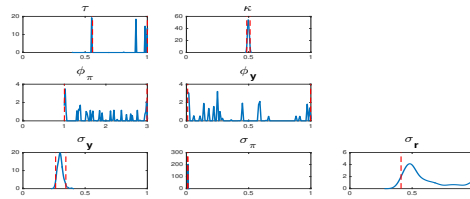
(d) RE model, US data (80)



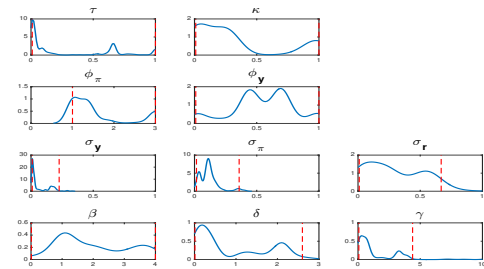
(e) BR model, US data (120)



(f) RE model, US data (120)



(g) BR model, US data (148)



(h) RE model, US data (148)

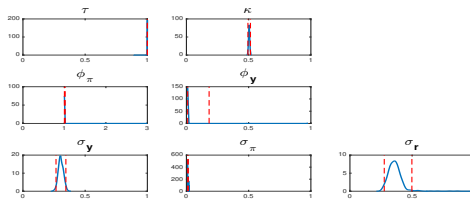


Figure 3: Densities of parameter estimates: forward-looking version. Blue curves depict kernel density estimates of the sample densities, and dashed red lines depict 95% intervals of sample estimates.

and specifications. For example, for the Great Moderation period with US(60), we have σ_y , σ_π and σ_r estimated at approximately or strictly zero in the BR NKM case, while σ_π is the only shock for which this is true for both parameters ϕ_y and ϕ_π in the RE NKM case.

It can be concluded from this observation that the data ‘push’ the point estimates to the pre-defined lower and upper bounds of the starting values. This implies that these boundaries have to be changed to probably avoid bias at all, e.g., allowing the value of the inherited persistence parameter τ to be greater than unity. Obviously, such a definition of the parameter space is at odds with the underlying economic and probability theory. Given the observation of bimodal-shaped parameter densities observed especially in BR NKM, the previous statement then raises the issue of model misspecification. As an example for the latter, based on all US subperiods in BR NKM, the dynamic IS equation and NKPC are decomposed from the remaining system of equations. This holds since the slopes of the Euler equation and the Phillips curve, τ and κ , are strictly estimated to be insignificant. Hence, inherited persistence in terms of cross-volatility within the inflation-output gap nexus is absent. Output and inflation dynamics are therefore only characterized by low cost-push shocks of degree 0.03 to 0.11 in this case. The same statements are true for the Euro Area but only with respect to κ , as τ is statistically significant.

In the following, we offer a brief economic interpretation of the parameter estimates in the BR NKM for the Euro Area dataset with 140 observations. We choose the latter since due to the large number of insignificant parameter estimates in the US cases, any discussion in this regard seems to be pointless. As we obtain $\tau = 0.68$, this indicates that the degree of inherited persistence in the output gap dynamics is moderated by applying the SMLE method. The type of persistence with respect to inflation dynamics cannot be interpreted since κ is estimated to be insignificant. The estimation result of $\phi_\pi = 2.14$ reveals a value greater than that implied by the Taylor principle. ϕ_y attains a comparatively low value of 0.78, which resembles a strong emphasis on inflation over output stabilization as expected by the European Central Bank. The standard deviations of the shock to inflation and the nominal interest rate exhibit estimation results of low magnitude given by $\sigma_\pi = 0.10$ and $\sigma_r = 0.08$, respectively. The remaining standard deviation σ_y is insignificant.

Note that all bounded rationality parameters β and δ are significant. The predicted subjective mean value of the output gap is estimated to be 1.42, which implies that optimistic agents were slightly optimistic over the same sample period. The opposite statement holds vice versa for the pessimistic agents due to the symmetric structure of the heuristics (5) and (6). This observation is justified by the fact that we have $\beta/2 = 0.71$. In economic terms, optimists (pessimists) believe that the future output gap will differ positively (negatively) by 0.71 percent on average from its steady-state value. We obtain a value for the divergence in belief parameter δ of 1.24. Again, due to symmetry, agents take any uncertainty (measured by the standard deviation $\lambda_{y,t}$) by the amount of $\delta/2 = 0.62$ regarding their subjective belief (β) into account. In comparison, for the US case, we observe significant values only for β given by 0.69 and 3.73 for US(60) and US(120), respectively. This implies that the forecast of output becomes more optimistic (pessimistic) about the predicted deviation from the steady state during the transition from the Great Inflation ($\delta/2 = 0.345$) to the Great Moderation ($\delta/2 = 1.865$) period.

A novelty of our empirical estimates is that we are able to pin down the intensity of choice param-

eter γ . The latter is estimated to be 7.01 for the Euro Area. In the US case, γ varies from 0.95 over 0.53 to 0.68 during the Great Inflation, Moderation, and Recession periods, respectively. All four of these estimates are statistically significant. This implies that the switching process becomes stochastic rather than deterministic, where the latter can be ruled out as judged by the confidence intervals. To extend this observation further, agents sort themselves in one of the different groups randomly a priori. Over time, they then have the incentive to employ their expectation formation process as applied before, i.e., in every period the switching from one forecast heuristic to the other is quite smooth and slow for all US subperiods but more intense for the Euro Area given a value of γ close to 10.¹⁰

5.4. Empirical Results: Hybrid Version of RE vs. BR

In the following section, we assess the extent to which the previous result holds when the hybrid version of the NKM is considered. There is broad agreement among macroeconomists that the hybrid version of the RE NKM is able to capture the inertia in the underlying economy while the purely forward-looking one cannot. This is no surprise because due to the RE hypothesis, forecasts of future development are zero in a non-persistent linear model with only white noise error terms being assumed. The attentive reader might already be aware that, thus far, we have not discussed the values of the average log-likelihoods (LL) of the different specifications presented in [Table 2](#) above. The reason for this is obvious: while both the BR NKM specifications exhibit a slightly better fit to the data than the RE NKM ones do, it is inaccurate to argue that the BR NKM outperforms the RE NKM in term of fitting. This becomes apparent when we recall that in the BR NKM with unconditional standard deviation ($\lambda_{y,t}$) and the heuristics of the inflation extrapolators (which is simply π_{t-1} ; cf. [equation \(12\)](#)), this accounts for the backward-looking components of the model. The RE NKM does not exhibit any backward-looking structure or autocorrelation in the shocks. This is crucial because the underlying time series for output, inflation and the nominal interest gaps are highly persistent.

[Table 3](#) and [Figure 4](#) report the results of the hybrid version of the model specification. It is immediately apparent that the values of the average log-likelihood are almost identical across specifications for both the Euro Area and US economy. This is in line with the overall findings in [Jang and Sacht \(2016\)](#). Those authors conclude that because of the almost identical values of their fitness criteria (i.e., the value of the objective function based on their SMM approach), the BR NKM is able to describe the data at least as well as the RE NKM. Interestingly, with respect to the significant parameter estimates of γ across all BR specifications, we observe a pairwise degree of switching on the almost identical magnitudes, as in the forward-looking case. This implies that the switching process remains rather stochastic but, again, to a greater extent in the Euro Area.

According to [Table 3](#), a number of the remaining parameters are, however, estimated to be insignificant. For the BR NKM, this is true for χ , α , κ , σ_y and δ in the Euro Area. For the US subperiods,

¹⁰Note that the results are comparable to those presented for the NKM under bounded rationality obtained via SMM in [Jang and Sacht \(2016, 2018\)](#), where a Taylor rule with smoothing is assumed. Note that the intensity of choice parameter is fixed in their study, which means that as a cross-check, the cases of $\gamma = 10$ for the Euro Area and $\gamma = \{0.1, 1\}$ for the US should be considered since we find $\gamma = 7.01$ for the Euro Area and $\gamma = \{0.95, 0.53, 0.68\}$ for US(60), US(120) and US(148), respectively.

Table 3: Results of the empirical study: hybrid version of RE vs. BR

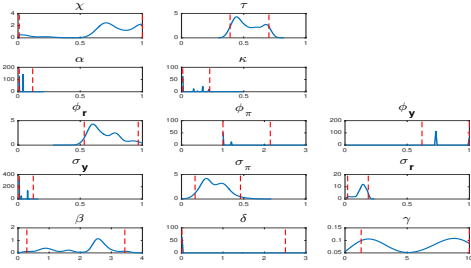
Par.	BR NKM				RE NKM			
	Euro (140)	US (60)	US (120)	US (148)	Euro (140)	US (60)	US (120)	US (148)
χ	.77	.42	.07	1	1	.91	.85	.79
$\langle 0, 1 \rangle$	(0-1)	(0-1)	(0-1)	(.38-1)	(1-1)	(.24-1)	(.73-1)	(.71-.95)
τ	.50	.39	.07	0	0	.17	0	0
$\langle 0, 1 \rangle$	(.39-.70)	(.04-1)	(.03-.32)	(0-.17)	(0-.02)	(0-.24)	(0-.02)	(0-.02)
α	0	.75	1	0	.05	.32	.14	.13
$\langle 0, 1 \rangle$	(0-.12)	(0-1)	(0-1)	(0-.17)	(0-.09)	(0-.82)	(0-.26)	(0-.37)
κ	0	.57	0	0	0	.08	.09	0
$\langle 0, 1 \rangle$	(0-.23)	(0-.71)	(0-.93)	(0-.14)	(0-.09)	(.04-.39)	(.03-.19)	(0-0)
ϕ_r	.66	.69	.88	.91	.50	.79	.74	.88
$\langle 0, 1 \rangle$	(.53-.97)	(0-.98)	(.81-.95)	(.85-.99)	(.45-.82)	(.63-.89)	(.73-.80)	(.84-.92)
ϕ_π	1	2.50	1.42	1	1	1.28	3	1
$\langle 1, 3 \rangle$	(1-2.14)	(1-3)	(1-2.77)	(1-3)	(1-1)	(1-3)	(1.41-3)	(1-1.31)
ϕ_y	1	1	1	.35	.79	.21	1	1
$\langle 0, 1 \rangle$	(.62-1)	(.16-1)	(0-1)	(0-1)	(.73-1)	(0-1)	(1-1)	(.31-1)
σ_y	0	0	.15	.24	.16	0	.14	.09
$\langle 0, 1 \rangle$	(0-.12)	(0-.06)	(.05-.21)	(.09-.34)	(.12-.18)	(0-.18)	(.11-.17)	(.06-.13)
σ_π	.25	.04	.19	.29	.18	0	.16	.33
$\langle 0, 1 \rangle$	(.11-.47)	(0-.47)	(.08-.34)	(.14-.44)	(.16-.42)	(0-.29)	(.08-.22)	(.23-.45)
σ_r	.14	.03	.16	.07	.05	0	.11	.10
$\langle 0, 1 \rangle$.02-.19	(0-.09)	(.08-.20)	(.03-.12)	(0-.16)	(0-.13)	(.08-.20)	(.07-.13)
β	2.53	2.62	3.61	1.90	-	-	-	-
$\langle 0, 4 \rangle$	(.28-3.43)	(0-4)	(.74-4)	(0-4)	-	-	-	-
δ	0	1.41	.37	0	-	-	-	-
$\langle 0, 3 \rangle$	(0-2.50)	(.22-3)	(0-2.78)	(0-.55)	-	-	-	-
γ	5.86	.92	.17	.52	-	-	-	-
$\langle 0, 10 \rangle$	1.38-10	(.05-10)	(.06-1.52)	(.08-1.36)	-	-	-	-
LL	.26	.37	.28	.30	.26	.26	.27	.30
	(.24-.27)	(.28-.44)	(.25-.30)	(.29-.32)	(.23-.26)	(.20-.34)	(.25-.27)	(.29-.31)

Constraints for optimization and initial conditions given in $\langle \rangle$ brackets. Sample medians reported, and 95% confidence intervals of sample estimates reported in $()$ parentheses. LL denotes the average log-likelihood of the estimated models and represents statistical fit. Figures rounded to 2 decimal places.

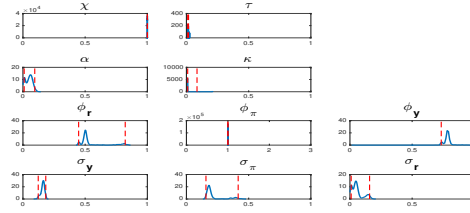
we have 8, 5 and 6 insignificant parameters for US(60), US(120) and US(148), respectively. By inspecting the RE NKM, we observe 7 insignificant parameters in the Great Inflation period, while only the estimates for τ and α are problematic in the Great Moderation and Recession periods. In addition, the parameter estimate of κ should also be mentioned in this regard for US(148) only. Estimation based on Euro Area data yields the largest number of insignificant parameters in this specification, namely, τ , α , κ and σ_r .

Regarding the parameters that measure inherited persistence, χ and α , the observations from our simulation study (not reported here) imply that the distributions of the corresponding pseudo-true parameters exhibit uniform and/or random shapes. As an example for the former, this true for χ as shown in Figure 4 (panel c), which represents the graphical outcome for the BR NKM based on US(60). This tendency does not seem to disappear rapidly enough given an increase in the number of observations. This puzzling result is interesting because in the purely-forward looking case, the pseudo-true parameters can be successfully estimated, especially given a large data set, without non-normal densities being detected. Potential explanations can be found at a theoretical and econometric level.

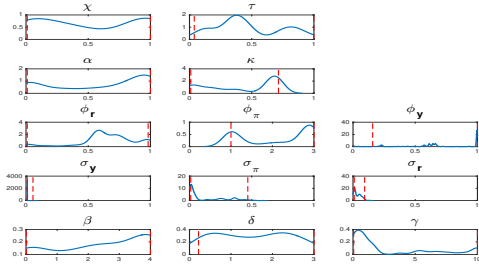
(a) BR model, Euro Area data (140)



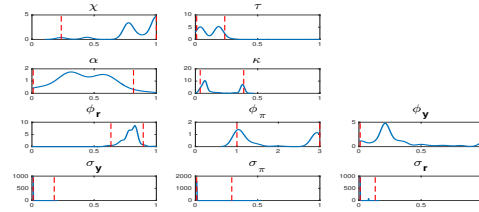
(b) RE model, Euro Area data (140)



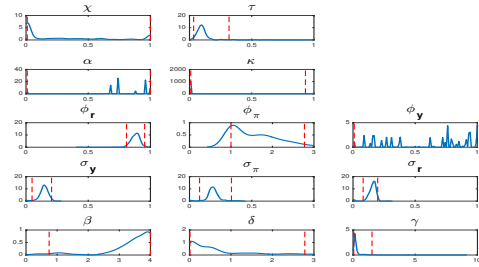
(c) BR model, US data (60)



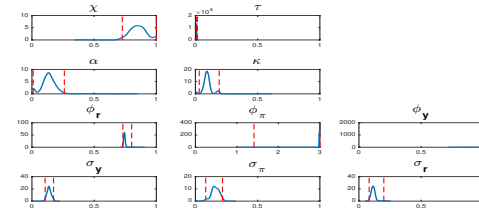
(d) RE model, US data (60)



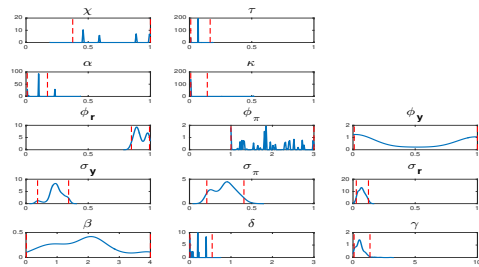
(e) BR model, US data (120)



(f) RE model, US data (120)



(g) BR model, US data (148)



(h) RE model, US data (148)

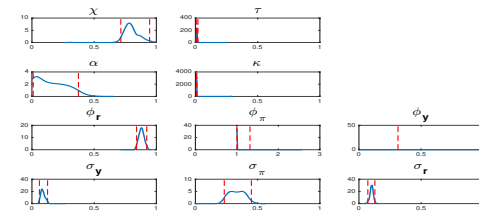


Figure 4: Densities of parameter estimates: hybrid version. Blue curves depict kernel density estimates of the sample densities, and dashed red lines depict 95% intervals of sample estimates.

First, the hybrid version of the NKM might simply suffer from model misspecification. In this case, alternative ways of introducing backward-looking components (other than habit formation in consumption and price indexation) into the model might be considered. A promising solution would be the use of forecast heuristics related to bounded rationality. According to our simulation study and as seen in [Table 2](#), this approach could indeed be effective based on a large data set, at least based on Euro Area data with 140 observations. This specification of the forward-looking model exhibits good performance in terms of simulation and estimation, namely, low numbers of insignificant and biased estimates. Second, forcing the SMLE procedure to optimize parameters that have a flat likelihood function (such as χ and α in our case) makes the estimates of the other parameters less precise. In this case, the overall performance of the estimation is dragged down if those ‘troublesome’ parameters are not identified a priori via a simulation exercise. In summary, we hesitate to offer an economic interpretation of the parameter estimates from the hybrid model. While a researcher might be tempted to consider these estimates given the better fit of this model version to the data compared to the forward-looking version, our observations must be interpreted as a warning against doing so.

6. Conclusion

In the absence of the rational expectation paradigm, a growing number of studies on dynamic stochastic equilibrium models address the importance of boundedly rational expectation formation. In this paper, we examine the baseline New-Keynesian model (NKM) with heterogeneous agents who adopt behavioral heuristics in forecasting future movements of output and inflation. In the model, agents exhibit an optimistic or pessimistic view about the future dynamics of output or act as inflation targeters or extrapolators when forming expectations based on discrete choice. The corresponding non-linear specifications of the model are for the first time estimated via the simulated maximum likelihood estimator (SMLE) approach ([Kristensen and Shin, 2012](#)) that considerably relaxes restrictive theoretical assumptions required competing estimation methods. We extend the univariate version of the SMLE applied in the financial econometrics literature to multivariate macroeconomic optimization problems and customize it for the behavioral NKM ([De Grauwe, 2011](#)).

In Monte Carlo simulations, we first analyze the ability of the SMLE to accurately estimate the pseudo-true parameter values of various competing specifications of the model. The most important novelty of this method is its capacity to estimate the pseudo-true intensity of choice under the discrete choice expectation updating scheme. In previous related studies, this parameter had to be calibrated or was estimated to be statistically insignificant and uninformative. On the contrary, our results show that the SMLE is clearly able to recover the pseudo-true switching parameter with surprising precision and with no consistent small sample bias. Moreover, another behavioral parameter of the model, the subjective mean of the future output gap, is estimated relatively precisely. In summary, our analysis suggests that by using the SMLE method, a majority of the 10 parameters of the behavioral NKM can be recovered reasonably well. Interestingly, the proposed method seems to favor estimation of the bounded rational (BR) model over the rational expectations (RE) model. The curse of dimensionality arises via a consistent downward bias in the estimates of the magnitude of idiosyncratic shocks across all simulations and does not tend to disappear with increasing sample size. Our study thus also reveals and stresses this consistent

imperfection that researchers need to consider carefully when interpreting empirical results.

Our main empirical results confirm that the forward-looking version of the model in both its BR and RE specifications exhibits good performance. This is especially true for the Euro Area. Based on the corresponding data set, which covers 140 observations, we report only insignificant estimates and up- and downward bias for a minority of parameters. We find evidence of a low degree of expected deviations of the output gap from its steady-state value from the perspectives of optimistic and pessimistic agents. A novelty of our analysis is that we are also able to pin down the parameter for the intensity of choice. The estimate reveals that the underlying switching process is stochastic rather than deterministic across all BR specifications and data sets. This accounts for the insensitive turnover from one forecast heuristic to the other over time. We argue that the empirical results for the hybrid version of the model must be treated with caution. Although the fit to the data indicates that the hybrid version performs better than the forward-looking version, the numbers of insignificant and biased parameters are larger in the hybrid version. Potential explanations include a misspecification of the model and the disadvantage of the SMLE procedure in coping with a flat likelihood function associated with specific parameters. Further exploration of this issue is needed.

We thus argue that this study considerably expands our understanding of the estimation of bounded rationality models with heterogeneous agents used in macroeconomic research. Specifically with respect to the intensity of choice parameter, our estimation results here are fruitful in terms of the parametrization of a bounded rationality model used for policy analysis. The focus of future work in this area should be on the estimation of much more complex macroeconomic models. Related to this, various types of different forecast heuristics might be added to those considered in this paper. Heuristics with backward-looking components are of particular interest to capture the observed high degree of inertia in macroeconomic data. We leave this to future research.

Acknowledgements

Kukacka gratefully acknowledges financial support from the UNCE project (UNCE/HUM/035). A significant part of the research leading to these results was conducted during the postdoctoral research stay by J. Kukacka at the University of California, Irvine, funded by the People Programme (Marie Curie Actions) of the European Unions Seventh Framework Programme FP7/2007-2013/ under REA grant agreement number 609642. Jang's work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A5A8018797).

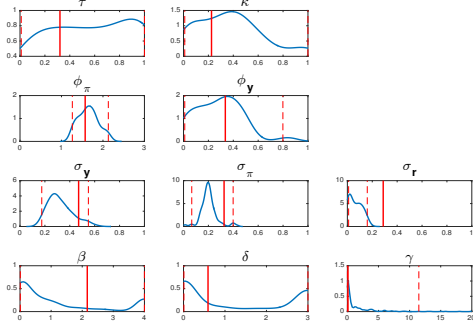
References

- Altissimo, F. and A. Mele (2009). Simulated non-parametric estimation of dynamic models. *The Review of Economic Studies* 76(2), 413–450.
- Andersen, T. G., L. Benzoni, and J. Lund (2002). An empirical investigation of continuous-time equity return models. *The Journal of Finance* 57(3), 1239–1284.
- Assenza, T., T. Bao, C. Hommes, and D. Massaro (2014). Experiments on expectations in macroeconomics and finance. In *Experiments in Macroeconomics*, Chapter 2, pp. 11–70. Emerald.
- Barde, S. and S. van der Hoog (2017). An empirical validation protocol for large-scale agent-based models. Technical report, Bielefeld Working Papers in Economics and Management No. 04-2017.
- Boswijk, H. P., C. H. Hommes, and S. Manzan (2007, February). Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics & Control* 31(2), 1938–1970.
- Brock, W. A. and C. H. Hommes (1997). A rational route to randomness. *Econometrica* 65(5), 1059–1095.
- Brock, W. A. and C. H. Hommes (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics & Control* 22, 1235–1274.
- Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2002). Can sticky price models generate volatile and persistent real exchange rates? *The Review of Economic Studies* 69(3), 533–563.
- Cox, J. C., J. Ingersoll, Jonathan E., and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica* 53(2), 385–407.
- Day, R. H. and M. Pingle (1991). Economizing economizing. In R. Franz, H. Singh, and J. Gerber (Eds.), *Handbook of Behavioral Economics*, pp. 509–22. Greenwich: JAI Press.
- De Grauwe, P. (2010). Top-down versus bottom-up macroeconomics. *CESifo Economic Studies* 56(4), 465–497.
- De Grauwe, P. (2011). Animal spirits and monetary policy. *Economic Theory* 47(2), 423–457.
- De Grauwe, P. (2012). Booms and busts in economic activity: A behavioral explanation. *Journal of Economic Behavior & Organization* 83(3), 484–501. The Great Recession: motivation for re-thinking paradigms in macroeconomic modeling.
- Deák, S., P. Levine, J. Pearlman, and B. Yang (2017). Internal rationality, learning and imperfect information. Technical report, University of Surrey.
- Doucet, A., N. J. Gordon, and V. Krishnamurthy (2001, Mar). Particle filters for state estimation of jump markov linear systems. *IEEE Transactions on Signal Processing* 49(3), 613–624.
- Fagan, G., J. Henry, and R. Mestre (2001). An area-wide model (awm) for the euro area. Technical Report Working Paper 42, European Central Bank.
- Fagiolo, G., A. Moneta, and P. Windrum (2007). A critical guide to empirical validation of agent-based models in economics: Methodologies, procedures, and open problems. *Computational Economics* 30(3), 195–226.
- Franke, R. (2009). Applying the method of simulated moments to estimate a small agent-based asset pricing model. *Journal of Empirical Finance* 16, 804–815.
- Franke, R. (2018). Competitive moment matching of a new-keynesian and an old-keynesian model. *Journal of Economic Interaction and Coordination* 13(2), 201–239.
- Franke, R., T.-S. Jang, and S. Sacht (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.
- Franke, R. and F. Westerhoff (2012). Structural stochastic volatility in asset pricing dynamics: Estimation and model contest. *Journal of Economic Dynamics and Control* 36(8), 1193–1211. Quantifying and Understanding Dysfunctions in Financial Markets.
- Franke, R. and F. Westerhoff (2018). Taking stock: A rigorous modelling of animal spirits in macroeconomics. *Journal of Economic Surveys* 31(5), 1152–1182.
- Gabaix, X. (2017, February 1). A behavioral new keynesian model. Working Paper Series 22954, National Bureau of Economic Research.
- Gaunersdorfer, A., C. H. Hommes, and F. O. O. Wagener (2008). Bifurcation routes to volatility clustering under evolutionary learning. *Journal of Economic Behavior & Organization* 67, 27–47.
- Grazzini, J. and M. Richiardi (2015). Estimation of ergodic agent-based models by simulated minimum distance. *Journal of Economic Dynamics & Control* 51, 148–165.
- Grazzini, J., M. G. Richiardi, and M. Tsionas (2017). Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control* 77, 26–47.
- Guerini, M. and A. Moneta (2017). A method for agent-based models validation. *Journal of Economic Dynamics and Control* 82, 125–141.

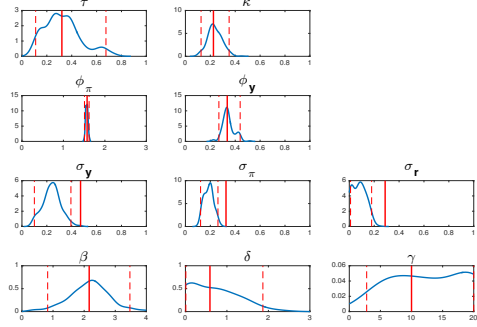
- Hommes, C. (2001). Financial markets as nonlinear adaptive evolutionary systems. *Quantitative Finance* 1(1), 149–167.
- Hommes, C. (2013). *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems*. Cambridge University Press. Cambridge Books Online.
- Hommes, C. H. (2006). Chapter 23: Heterogeneous agent models in economics and finance. In L. Tesfatsion and K. Judd (Eds.), *Handbook of Computational Economics*, Volume 2 of *Handbook of Computational Economics*, pp. 1109–1186. Elsevier.
- Jang, T.-S. and S. Sacht (2016). Animal spirits and the business cycle: Empirical evidence from moment matching. *Metroeconomica* 67(1), 76–113.
- Jang, T.-S. and S. Sacht (2017). Modelling consumer confidence and its role for expectation formation: A horse race. Technical report, Kiel University WP No2017-04.
- Jang, T.-S. and S. Sacht (2018). Forecast heuristics, consumer expectations, and new-keynesian macroeconomics: A horse race. Technical report, Kiel University WP No2018-09.
- Kristensen, D. (2009, 10). Uniform convergence rates of kernel estimators with heterogeneous dependent data. *Econometric Theory* 25, 1433–1445.
- Kristensen, D. and Y. Shin (2012). Estimation of dynamic models with nonparametric simulated maximum likelihood. *Journal of Econometrics* 167(1), 76–94.
- Kukacka, J. and J. Barunik (2017). Estimation of financial agent-based models with simulated maximum likelihood. *Journal of Economic Dynamics & Control* 85, 21–45.
- Lamperti, F. (2018). An information theoretic criterion for empirical validation of simulation models. *Econometrics and Statistics* 5, 83–106.
- Lamperti, F., A. Roventini, and A. Sani (2018). Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control* 90, 366–389.
- LeBaron, B. and L. Tesfatsion (2008). Modeling macroeconomics as open-ended dynamic systems of interacting agents. *The American Economic Review* 98(2), 246–250.
- Lee, D. and K. Song (2015). Simulated maximum likelihood estimation for discrete choices using transformed simulated frequencies. *Journal of Econometrics* 187(1), 131–153.
- Liu, C. and P. Minford (2014). Comparing behavioural and rational expectations for the us post-war economy. *Economic Modelling* 43, 407–415.
- Lux, T. (2018). Estimation of agent-based models using sequential monte carlo methods. *Journal of Economic Dynamics and Control* 91, 391–408. Special Issue in Honour of Prof. Carl Chiarella.
- Lux, T. and R. C. Zwickels (2018). Chapter 8 - empirical validation of agent-based models. In C. Hommes and B. LeBaron (Eds.), *Handbook of Computational Economics*, Volume 4 of *Handbook of Computational Economics*, pp. 437–488. Elsevier.
- Milani, F. (2007). Expectations, learning and macroeconomic persistence. *Journal of Monetary Economics* 54(7), 2065–2082.
- Munier, B., R. Selten, D. Bouyssou, P. Bourguine, R. Day, N. Harvey, D. Hilton, M. J. Machina, P. Parker, J. Sterman, E. Weber, B. Wernerfelt, and R. Wensley (1999, Aug). Bounded rationality modeling. *Marketing Letters* 10(3), 233–248.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- Windrum, P., G. Fagiolo, and A. Moneta (2007). Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation* 10(2), 8.

Appendix

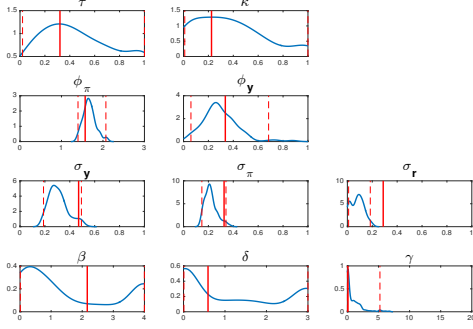
(a) $\gamma = 0.1, T = 100$



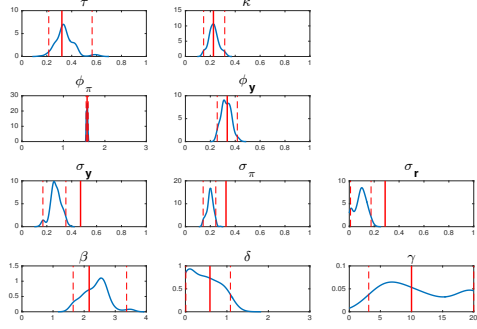
(b) $\gamma = 10, T = 100$



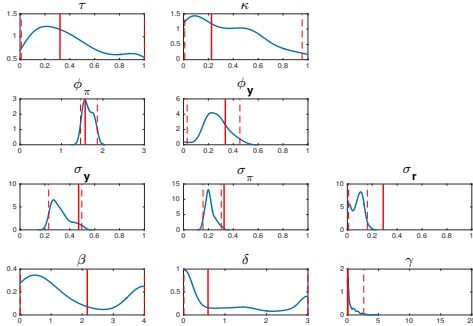
(c) $\gamma = 0.1, T = 200$



(d) $\gamma = 10, T = 200$



(e) $\gamma = 0.1, T = 400$



(f) $\gamma = 10, T = 400$

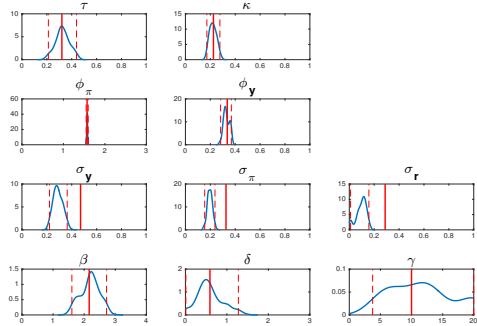


Figure 5: Densities of the pseudo-true parameter estimates for the BR model with $\gamma = \{0.1, 10\}$. Blue curves depict kernel density estimates of the sample densities, bold red lines show the pseudo-true values, and dashed red lines depict 95% intervals of sample estimates.