

Kukacka, Jiri; Sacht, Stephen

Working Paper

Estimation of Heuristic Switching in Behavioral Macroeconomic Models

Economics Working Paper, No. 2021-01

Provided in Cooperation with:

Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Kukacka, Jiri; Sacht, Stephen (2021) : Estimation of Heuristic Switching in Behavioral Macroeconomic Models, Economics Working Paper, No. 2021-01, Kiel University, Department of Economics, Kiel

This Version is available at:

<https://hdl.handle.net/10419/231265>

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.

Estimation of Heuristic Switching in Behavioral Macroeconomic Models

by Jiri Kukacka and Stephen Sacht



Estimation of Heuristic Switching in Behavioral Macroeconomic Models*

February 11, 2021

Jiri Kukacka^{a,b,*}, Stephen Sacht^c

^a*Czech Academy of Sciences, Institute of Information Theory and Automation, Czechia*

^b*Charles University, Faculty of Social Sciences, Institute of Economic Studies, Czechia*

^c*Kiel University, Department of Economics, Germany*

Abstract

This paper offers a simulation-based method for the estimation of heuristic switching in nonlinear macroeconomic models. Heuristic switching is an important feature of modeling strategy since it uses simple decision rules of boundedly rational heterogeneous agents. The simulation study shows that the proposed simulated maximum likelihood method identifies the behavioral effects that stay hidden for standard econometric approaches. In the empirical application, we estimate the structural and behavioral parameters of the US economy. We are especially able to reliably identify the intensity of choice that governs the models' nonlinear dynamics.

JEL: C53, D83, E12, E32.

Keywords: Behavioral Heuristics, Heuristic Switching Model, Intensity of Choice, Simulated Maximum Likelihood.

*This paper benefited greatly from Reiner Franke's and Mark Weder's detailed comments and research discussions with William Branch during Kukacka's stay at the University of California, Irvine. We are also indebted to the participants of WEHIA 2017, CEF 2018, the Second Bamberg Behavioral Macroeconomics Workshop 2019, the EAEPE Conference 2019, CFE 2019, the Annual Conference of the Eastern Economic Association 2020, and the EAEPE Conference 2020 for their many useful comments and suggestions. Very special thanks go to Marco Del Negro for his challenging support. We also thank Tae-Seok Jang for his help with the MLE and Jozef Barunik for his proposals for final edits.

Kukacka gratefully acknowledges the financial support from the Czech Science Foundation under the 20-14817S "Linking financial and economic agent-based models: An econometric approach" project, from the Charles University PRIMUS program [project PRIMUS/19/HUM/17], and from the Charles University UNCE program [project UNCE/HUM/035].

*Correspondence to: Jiri Kukacka, Pod Vodarenskou vezi 4, 182 00 Prague 8, Czechia, +420 602 767 305

Email addresses: jiri.kukacka@fsv.cuni.cz (Jiri Kukacka), sacht@economics.uni-kiel.de (Stephen Sacht)

1. Introduction

The implications of the departures from fully rational agents have become one of the central issues of macroeconomic modeling. As suggested by [Simon \(1972\)](#), economic agents act under bounded rationality (BR) instead of perfectly processing all information using unlimited cognitive capacities. In most situations, human behavior follows simple decision rules known as behavioral heuristics that have proven to be suitable in the past.

The theoretical literature has been attempting for a long time to formalize deviations from rational expectations (RE) and to describe the decision making of boundedly rational heterogeneous agents. One such possible macroeconomic framework is business cycle models with heuristics and a nonlinear switching mechanism ([De Grauwe, 2010](#); [Hommes et al., 2019](#)). However, these approaches have not yet been satisfactorily reflected in the empirical literature, and it is often difficult to identify some behavioral effects in the current nonlinear macroeconomic models. We thus propose a new econometric method that allows estimating these models, especially to identify important behavioral parameters.

Although the related econometric research has made significant progress in the last decade, using these types of models with the empirical data is still challenging. First, the choice of the empirical variables is not straightforward, as standard econometric tools assume the stationarity of the input data that is often violated for macroeconomic time series. Second, because of a relatively large number of unknown parameters, some of the coefficients might not be identified. This requires new tools for parameter estimation ([Del Negro et al., 2007](#); [Kleibergen and Mavroeidis, 2014](#); [Kulish and Pagan, 2017](#)). Third, although nonlinear models can be represented in the state-space form, analytical solutions of the joint probability density seldom exist to allow for, e.g., the maximum likelihood inference. Fourth, because of a nonlinear structure, “a possibly non-monotonic likelihood surface . . . tends to be rugged making it challenging to find a global optimum” ([Lux and Zwinkels, 2018](#)). Finally, some authors emphasize the insufficient testing of model ergodicity ([Grazzini and Richiardi, 2015](#); [Fagiolo et al., 2019](#)); however, a toolkit for ergodicity testing for dynamic nonlinear multivariate models is still extremely scarce. The presented challenges and rapidly increasing operational capacity of computers have motivated simulation-based estimation methods that make

empirical inference possible for analytically intractable models.

We propose a simulated maximum likelihood (SML) estimation method, which addresses the first four issues mentioned above. Our main contribution lies in the modification of the univariate SML estimator and its transfer from financial econometrics ([Altissimo and Mele, 2009](#); [Kristensen and Shin, 2012](#); [Lee and Song, 2015](#); [Kukacka and Barunik, 2017](#)) to macroeconomic optimization. We first pursue a simulation study to test its performance in a controlled environment. Compared to the standard maximum likelihood estimator (MLE), the SML technique numerically approximates each observation’s conditional density via a standard kernel method. The derivation of the SML estimator is then similar to the MLE. Through a kernel approximation, the method elegantly bypasses the distributional assumptions of the MLE, the Bayesian approach or the moment selection problem of the simulated method of moments (SMM). It also leads to a smooth surface of the simulated likelihood function, which generally supports the optimization search. Finally, we illustrate the first empirical application of the multivariate SML for the estimation of a heuristic switching New-Keynesian model (NKM).

The model specification relates to the research on alternative explanations of expectation formation where agents apply simple forecasting heuristics. Agents form their individual expectations in a boundedly rational way based on past experience and dynamically sort themselves into groups populated by individuals who apply a particular heuristic. The heuristics are microfounded based on the previous literature on laboratory experiments and allow for heterogeneity in different groups. A suitable mechanism for such a dynamic endogenous sorting of agents is the multinomial logistic switching approach. This approach has become a prominent feature of models in financial economics over the last two decades ([Hommes, 2006](#); [Franke and Westerhoff, 2012](#); [Dieci and He, 2018](#), among others) and recently, also in system-of-equations models such as the dynamic stochastic general equilibrium (DSGE)-type. We recognize [Hommes et al. \(2019\)](#) and [De Grauwe and Ji \(2020\)](#) as the most recent key contributions to DSGE modeling under BR, and we directly build on their work.

As an empirical novelty, we report the statistically significant estimates for the macroeconomic intensity of choice, which is a parameter that governs the multinomial logistic switching

mechanism. The intensity of choice drives the agents’ sensitivity regarding the individual heuristics, and its value is crucial for the switching process from one group to another. Several previous macroeconomic studies have had difficulties estimating this parameter as heuristic switching models exhibit a much more complex nonlinear structure compared to linearized DSGE models under RE ([Jang and Sacht, 2016](#); [Lux and Zwickels, 2018](#)). Using simulations, we first show that the SML method can identify this parameter. We then successfully identify the empirical value of the intensity of choice for the US economy that is approximately between 1.4 and 1.5. We thus find the switching process to be less intense and rather smooth. We also estimate all additional behavioral parameters together with structural ones, especially the Taylor rule parameters. Thus, our results help to understand the implications of shocks and policy interventions in the absence of RE.

This paper proceeds as follows. The next section provides an overview of the related literature. [Section 3](#) describes the heuristic switching NKM. In [Section 4](#), we introduce the SML estimation method modified for multivariate macroeconomic models. [Section 5](#) presents a Monte Carlo simulation study of the suggested estimation framework’s performance before we discuss the empirical results in [Section 6](#). Finally, [Section 7](#) concludes the paper. The technical details and additional results are relegated to a regular Appendix and an [Online Appendix](#).

2. Related Literature

The analysis of behavioral macroeconomic models has become prominent over the last decade; see [Franke and Westerhoff \(2017\)](#) for an overview. The reasons are manifold. For example, [Hendry and Mizon \(2010\)](#) state that conditional rational expectations are not unaffected by structural breaks in economic times series. As a result, neither unbiased nor minimum mean-squared error predictions are observed. The authors call for a state-contingent expectation formation process that should be flexible within the standard prediction methods. [Coibion and Gorodnichenko \(2012, 2015\)](#) argue that information rigidities lead to a delayed response of mean forecasts to macroeconomics shocks. They show that a model framework with noisy information processing is much more empirically sound in displaying the expectation formation capability of agents compared to a model under RE.

Various approaches also advocate the design of optimal policy strategies within a model framework under BR. Related to this development, macroeconomists face the question of how

monetary and fiscal policy interventions can be conducted in the presence of non-RE formation. [Angeletos and Lian \(2018\)](#) discuss agents’ response to news about future changes in policy interest rates. The authors emphasize that heterogeneous beliefs avoid the so-called “forward guidance puzzle” ([Giannoni et al., 2015](#)). If a central bank promises to hold interest rates lower in the future, then a standard DSGE model under RE overestimates the impact of forward guidance on the economy. [Woodford \(2019\)](#) suggests that an objective function related to the outcome being achieved under forward guidance should be considered, which might solve the puzzle. As an essential feature of his approach, the objective function is updated under non-RE formation techniques such as constant gain learning. [Goy et al. \(2020\)](#) allow for an endogenous expectation formation process of private agents concerning central bank credibility and study the effectiveness of forward guidance in a switching model similar to the one discussed in our paper.

2.1. Heuristic Switching Models in Macroeconomics

Although heuristic switching models have been considered in financial economics for more than two decades ([Brock and Hommes, 1997](#)), in macroeconomics, such a modelling approach is rather new ([De Grauwe, 2010, 2011](#)). Heuristics represent simple rules of behavior that originate from the fact that the structure of the economy is observable, but the interactions between relevant variables such as output and inflation are barely comprehensible ([Munier et al., 1999](#)). Such boundedly rational behavior in response to a lack of complete information is based on habits, imitation and/or procedural optimization ([Day and Pingle, 1991](#)). Related research questions have already drawn interest in the field of experimental economics; see, e.g., [Anufriev et al. \(2016, 2018\)](#). Various studies reveal that switching between forecasting heuristics based on the multinomial logistic model can indeed be observed within a laboratory environment; we refer to [Assenza et al. \(2014\)](#) and [Hommes \(2021\)](#) for literature reviews that include corresponding evidence in a DSGE context.

The importance of studying these models seems to be undeniable. [Haldane and Madouros \(2012\)](#) emphasize the “less-is-more” principle, whereby the cost of cognition and information processing might outweigh the desired optimized response in a complex environment. [Gigerenzer and Brighton \(2009\)](#) empirically show that a biased mind can achieve much more accurate results regarding desired preferences when using heuristics. Therefore, resource-intensive forecasting

strategies might be subconsciously avoided due to humans’ limited cognitive abilities although they may increase forecast accuracy in a complex environment.

Two very recent contributions that we directly build on are [De Grauwe and Ji \(2020\)](#) and [Hommes et al. \(2019\)](#). [De Grauwe and Ji \(2020\)](#) study the effects of structural reforms regarding price flexibility and labor market rigidities in a hybrid heuristic switching NKM, i.e., a model framework with leads and lags. The authors show that their model accounts for the empirical observation of a non-Gaussian distributed and highly persistent output gap. [Hommes et al. \(2019\)](#) consider a forward-looking model analogical to the model of [De Grauwe and Ji \(2020\)](#) but with a higher number of heuristics to choose from. The microfoundations of the corresponding heuristics based on experiments are presented in [Anufriev and Hommes \(2012\)](#). The authors provide evidence that the heuristic switching macroeconomic models can fit the empirical data obtained in a laboratory environment very well.

2.2. Estimation of Heuristic Switching Models

The empirical estimation of behavioral models has rapidly gained attention in macroeconomics over the years, which naturally include heuristic switching models. Before specific examples are provided, it might be important to distinguish between estimation and two related concepts of applying economic models to data, specifically, validation and calibration. By validation, we specifically mean the empirical validation of model outputs that [Fagiolo et al. \(2019\)](#) aptly define as a general “process of evaluating the extent to which the outcome of a simulated model is a good representation of real-world observations”.¹ This is achieved in practice through Monte Carlo simulations that allow for verification to the extent that the model can replicate the characteristic patterns and regularities of real-world datasets, e.g., a set of stylized facts or moments of distributions of various economic variables. Both empirical and simulated data are thus essential for a validation analysis.

Model estimation, in contrast, is typically performed purely with empirical data. The aim is to assess the approximate size of an economic phenomenon. Estimation is a traditional rigorous

¹The growing interest in behavioral macroeconomic models has recently been accompanied by novel approaches to their empirical validation ([Guerini and Moneta, 2017](#); [Barde and van der Hoog, 2017](#); [Lamperti, 2018b](#); [Lamperti et al., 2018](#); [Barde, 2017, 2020](#)). [Lux and Zwinkels \(2018\)](#) provide an excellent survey of these attempts.

optimization approach for which a loss function is unambiguously defined. The estimated parameter value then provides the best fit in a clearly defined statistical sense. Finally, model calibration is generally understood as a process of “tuning” a model, i.e., the careful manipulation of its parameters until the model accurately mimics pre-defined characteristics. Compared to estimation, model calibration in its simplest form can be conducted even without any empirical data and subsequent iterations similar to a model parametrization according to some best practices in the literature. As previous econometric estimates often provide a basis for model calibration and because both are often a part of a more general validation analysis, the difference can become fuzzy. Moreover, there is a considerable terminological overlap that leads to some confusion in the literature as all three concepts largely use similar tools. [Kukacka \(2019\)](#) therefore compares and contrasts these concepts in detail by focusing on their main advantages and potential drawbacks.

Among the first to estimate a heuristic switching macroeconomic model are [Liu and Minford \(2014\)](#) who employ the indirect inference method ([Gourieroux et al., 1993](#)). The authors primarily test the model by [De Grauwe \(2010\)](#) against US data and strongly reject the behavioral model in favor of the RE model. However, a set of numerical parameter values via indirect estimation is also provided with a low estimated value (0.85) for the intensity of choice being reported. The estimates’ standard errors are not available, which calls into question the statistical significance of the true parameters. [Grazzini et al. \(2017\)](#) estimate the model by [De Grauwe \(2012\)](#) via Bayesian techniques. They report a limited effect of the intensity of choice on the model behavior as the corresponding posterior distribution does not depart significantly from the prior one. Other proponents of the Bayesian estimation of a heuristic switching NKM are [Deák et al. \(2017\)](#). [Jang and Sacht \(2016, 2019\)](#) apply the SMM but leave the intensity of choice unestimated for the [De Grauwe \(2011\)](#) model. Instead, they follow a grid calibration strategy in which they compare the results while varying the intensity of choice between 0.1 and 100.

2.3. Advantages and Limitations of the Simulated Maximum Likelihood Method

The SML estimator addresses most of the potential issues related to the abovementioned estimation methods. It is universally applicable without additional theoretical assumptions, it does not require the stationarity of the input variables, and it further reduces the discretionary choices necessary

for practical implementation. Compared to the indirect estimation, it requires neither an auxiliary time-series model nor the initial parametrization to be selected. Compared to the SMM, it avoids the selection of moments and the weighting matrix, which is to some extent inevitably “arbitrary and different choices may lead to differing estimation results” (Fagiolo et al., 2019). Finally, compared to the Bayesian approach, it reduces a potential arbitrariness of prior distributions’ specification. One can reliably take advantage of the Bayesian inference especially if the prior distributions of the macroeconomic variables are generally accepted. Of course, Bayesian methods can also be used with uninformative uniform distributions if there is no consensus about prior distributions for some estimated parameters. SML also naturally allows for empirical discrimination among alternative models similarly to the MLE, while “Bayesian methods cannot be used to test models against the data” (Liu and Minford, 2014) based on the traditional hypothesis testing procedure.

Perhaps most importantly, the kernel approximation generally leads to a smooth surface of the simulated likelihood function, which supports an optimization search. This is in sharp contrast especially with the SMM for which an often rugged surface of the objective function “would render standard derivative-based optimization routines useless” (Lux and Zwinkels, 2018). The SML approach also tends to be more asymptotically efficient as the entire empirical distribution is taken into account, not just selected moments. Finally, SML allows for the estimation of the intensity of choice that governs the switching mechanisms in the behavioral macroeconomic models that was found to be difficult to estimate by related studies that employed other inference methods in the past (Lux and Zwinkels, 2018).

The main limitation shared with all methods above lies in a relatively high computational burden. Another disadvantage of the SML approach is that the standard deviations of the idiosyncratic independent and identically distributed (i.i.d.) shocks have to be parametrized as the method requires their distribution to be known.

Another promising approach to some extent similar to SML is designed especially for the state space models and is based on recent advances in the Markov Chain Monte Carlo estimation technique (Doucet et al., 2015; Herbst and Schorfheide, 2015; Fernández-Villaverde et al., 2016;

[Fernández-Villaverde and Guerrón-Quintana, 2020](#)). In the so-called sequential Monte Carlo, which is also known as a particle filter, the distribution of an unobserved variable can be approximated by a swarm of particles propagated from one time step to the next via a resampling algorithm ([Doucet et al., 2001](#); [Fernández-Villaverde and Rubio-Ramírez, 2007](#); [Herbst and Schorfheide, 2014](#); [Lux, 2018](#)). However, we are not aware of applying the particle filter to the estimation of heuristic switching macroeconomic models. This may be because the resulting likelihood approximation is not a smooth function of the parameters ([Lux and Zwinkels, 2018](#)), which generally complicates an optimization search, and also because computational costs can become extensive.

3. Expectation Formation in the Heuristic Switching Model

3.1. The Core Structure of the Baseline NKM

The baseline NKM represents a realistic macroeconomic model with multivariate observables. It allows the study of the performance of the SML estimator in a more complex framework compared to univariate applications in finance.² 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 superscripts $j = \{\text{BR}, \text{RE}\}$ refer to the BR and the RE model specification, respectively. Therefore, we distinguish between the BR NKM and the hybrid RE NKM. The corresponding expectations operator is \tilde{E}_t^j , which is explicitly specified below for both models.

All variables are given in quarterly magnitudes. In equation (1), the dynamic IS curve results from the intertemporal optimization of consumption and saving, which leads to consumption smoothing. Parameter $\tau \geq 0$ denotes the inverse intertemporal elasticity of substitution in consumption behavior. Equation (2) represents the New-Keynesian Phillips Curve, where the output

²Examples are given by [Kristensen and Shin \(2012\)](#) and [Kukacka and Barunik \(2017\)](#). [Kristensen and Shin \(2012\)](#) estimate the following three types of models: the short-term interest rate model by [Cox et al. \(1985\)](#); the jump-diffusion model of daily stock returns by [Andersen et al. \(2002\)](#); and generic Markov decision processes. [Kukacka and Barunik \(2017\)](#) estimate the [Brock and Hommes \(1998\)](#) heterogeneous agent model. An introduction to the baseline NKM is presented in [Galí \(2015\)](#).

gap (y_t) acts as the driving force for inflation dynamics that originates from monopolistic competition and Calvo-type sticky prices. The slope of the New-Keynesian Phillips Curve is given by parameter $\kappa \geq 0$. Parameter ν represents the discount factor ($0 < \nu < 1$). Intrinsic persistence is incorporated into the demand and supply equations 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 $\varepsilon_{\{y,\pi,r\},t}$, which are i.i.d. with mean zero and variances $\sigma_{\{y,\pi,r\}}^2$.

3.2. The Model under Bounded Rationality

For the BR specification of the model, we directly follow the modelling approach by [Hommes et al. \(2019\)](#) where in its core, the purely forward-looking version of the baseline NKM is considered, i.e., with the intrinsic persistence parameters χ , α , and ϕ_r all set to zero. Regarding the expectation formation process, specific heuristics are applied, which are discussed below. The empirical microfoundation of these heuristics is discussed in [Anufriev and Hommes \(2012\)](#) who conduct learning-to-forecast experiments given a standard asset pricing model taken from financial economics. The authors report that the heuristics being used account for the stylized facts of a slow (almost) monotonic convergence, persistent oscillations with almost constant amplitude, and/or large initial but then dampening oscillations observed in the dynamics of their model. [Hommes et al. \(2019\)](#) confirm these findings based on their own experimental macroeconomic set-up where all three types of heuristics are supported by experimental evidence.

The following heuristics are considered in [Hommes et al. \(2019\)](#) and described in [Anufriev and Hommes \(2012\)](#):

$$E^{ADA}x_{t+1} = \eta x_{t-1} + (1 - \eta)x_t, \quad (4)$$

$$E^{TR}x_{t+1} = x_{t-1} + \iota(x_{t-1} - x_{t-2}), \quad (5)$$

$$E^{LAA}x_{t+1} = \mu(x_{t-1}^{obs} - x_{t-1}) + (x_{t-1} - x_{t-2}). \quad (6)$$

Similarly to [Hommes et al. \(2019\)](#), we apply these heuristics with respect to both the output gap

and inflation rate expectations, which implies that $x = \{y, \pi\}$. Agents therefore sort themselves into three groups of output gap and inflation forecasters. Based on the *adaptive* (*ADA*) heuristic (4), future expectation results from the weighted sum of the previous and current realizations of x with $0 \leq \eta \leq 1$. In the polar case $\eta = 1$, the expression stands for a static/naïve expectation formation process. In the *trend-following* (*TR*) heuristic (5), the past realization is considered while this forecasting rule follows the direction of the last change in x given by the term $(x_{t-1} - x_{t-2})$. The corresponding parameter of extrapolation denoted by $\iota \geq 0$ is crucial for observing specific patterns in the variable's dynamic. Although small values of ι around a value of 0.5 give rise to rather monotonic convergent movements, relatively large values already around unity might induce convergent but oscillatory dynamics. Finally, in the *learning anchoring and adjustment* (*LAA*) heuristic, the last change in x given by $(x_{t-1} - x_{t-2})$ is further extrapolated from an anchor learned through a sample average of the previous realizations of x denoted by x_{t-1}^{obs} . The corresponding extrapolation parameter is given by $\mu \geq 0$.

For the BR specification, the switching from one group to the other is based on the multinomial logistic model. The expression for the market forecast regarding the output gap and the inflation rate across the three groups is given by

$$\tilde{E}_t^{BR} x_{t+1} = \sum_{i=1}^3 (\alpha_{x,t}^{k\{i\}} \cdot E_t^{k\{i\}} x_{t+1}), \quad (7)$$

with $k = \{ADA, TR, LAA\}$. The probability $\alpha_{x,t}^k$ represents stochastic behavior by the agents who adopt a particular forecasting heuristic. Thus, $\alpha_{x,t}^k$ can be interpreted as the probability of being an adaptive forecaster, a trend-follower, or an anchored forecaster with respect to the development of the output gap and inflation rate at time t . The selection of the forecasting heuristics (4)-(6) depends on the forecast performance of each group U_t^k , which is given by the mean squared forecasting error. The utility of the forecast performance can be simply updated in every period (Brock and Hommes, 1997) as

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

where the memory parameter ρ (with $0 \leq \rho \leq 1$) determines the speed of a geometric dilution

of the impact of the lags of $U_{x,t}^k$ on the current utility of the forecast performance. Here, $\rho = 0$ suggests that agents do not consider the past observations of $U_{x,t}^k$ in their updating scheme (8) at all, while $\rho = 1$ means that they give the same weight of 1 to all past observations. The switching mechanism is that agents can adaptively revise their expectations given the forecast performance of particular heuristics based on the multinomial logistic model:

$$\alpha_{x,t}^k = \frac{\exp(\gamma U_t^k)}{\sum_{i=1}^3 \exp(\gamma U_t^{k\{i\}})}, \quad (9)$$

where the parameter $\gamma \geq 0$ denotes the intensity of choice, a parameter crucial to the stability of the system (cf. [Hommes, 2013](#); [Jang and Sacht, 2016](#), among others). Note that a negative γ lacks economic sense as it would imply irrational switching towards less precise forecasting heuristics.

Although we consider the BR model by [Hommes et al. \(2019\)](#) as our flagship one due to the existing micro evidence for the applied heuristics, in Appendix B, we conduct an additional Monte Carlo exercise similar to [Section 5](#) for various different switching models that are directly taken from the macroeconomic literature. Our focus is on the estimation of models with alternative sets of heuristics since transferring the SML technique from financial economics to macroeconomics is the main goal of our paper. We believe this additional robustness analysis markedly fosters our intentions in this regard and enables us to forestall potential criticism related to the “wilderness of bounded rationality” argument by [Sims \(1980\)](#).

3.3. The Model under Rational Expectations

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

$$\tilde{E}_t^{RE} x_{t+1} = E_t x_{t+1}, \quad (10)$$

with, again, $x = \{y, \pi\}$, and where E_t denotes the statistical expectation operator conditional on information at time t . For the corresponding random error term usually to be found in equation (10) and denoted by $\tilde{\varepsilon}_{x,t}$, $E_t \tilde{\varepsilon}_{x,t} = 0$ holds since it is independent of the future realizations in x . This implies that agents’ expectations are not systematically biased under RE.

In the Monte Carlo study and for the empirical application, we follow [Jang and Sacht \(2016\)](#)

and consider the BR NKM and the baseline RE NKM in the hybrid version with a lead and lag structure. With respect to the former, $\chi = \alpha = \phi_r = 0$ holds. We distinguish between both specifications according to a potential “persistence anomaly” (Chari et al., 2002), where the DSGE models that produce only monotonic dynamic patterns fail to empirically capture observable hump-shaped movements in output and inflation. This is indeed true for a purely forward-looking RE model, i.e., without any lag terms in the demand and supply equations (1) and (2), especially in the absence of auto-correlated shocks. However, in the BR model framework, the inertia in the model variables is ensured based on the heuristics (4)-(6) being applied. Being aware of the persistence in real-world data, we therefore explicitly account for intrinsic persistence due to the assumption of consumption habits, price indexation, and interest rate smoothing in the hybrid baseline NKM under RE. This explains why the term “hybrid” is attached to the RE NKM expression as it helps to clearly distinguish between both model specifications.

3.4. Model Solutions and Estimation

The state space representation of the baseline NKM is given by

$$\mathbf{A}\mathbf{X}_t + \mathbf{B}\mathbf{X}_{t+1}^j + \mathbf{C}\mathbf{X}_{t-1} + \mathbf{D}\mathbf{\Gamma}_t = \mathbf{0}, \quad (11)$$

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

$$\mathbf{X}_t = -\mathbf{A}^{-1}[\mathbf{B}\mathbf{X}_{t+1}^j + \mathbf{C}\mathbf{X}_{t-1} + \mathbf{D}\mathbf{\Gamma}_t] \quad (12)$$

is then obtained by applying the method of undetermined coefficients. For the BR specification, the forward-looking elements in X_{t+1}^{BR} are replaced by the forecasting heuristics (4)-(6), obviously except for the expectations on the interest rate $\tilde{E}_t^j r_{t+1}$. Since the BR NKM exhibits a backward-looking structure due to the heuristics (4)-(6), it has to be solved by backward induction. The solution for the hybrid RE NKM is instead obtained by applying the brute force iteration method introduced by Binder and Pesaran (1999).

With respect to the BR model specification, in the following section, we describe how the SML method estimates the *structural* $(\tau, \kappa, \phi_y, \phi_\pi)$ and the *bounded rationality* $(\eta, \iota, \mu, \gamma)$ param-

eters. For the RE model, the parameters for habit formation, price indexation, and interest rate smoothing (χ , α , and ϕ_r) are estimated, while the BR parameters are not considered at all.

4. The Simulated Maximum Likelihood Approach for the NKM

This section introduces the SML estimator to macroeconometrics. The SML method is primarily known from the financial econometric literature on the estimation of univariate time series models (Kristensen and Shin, 2012; Lee and Song, 2015; Kukacka and Barunik, 2017). For macroeconomic optimization problems, it requires subsequent modification to a multivariate version.

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:

$$z_t = m_t(x_t, \varepsilon_t, \theta), \quad t = 1, \dots, T, \quad (13)$$

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

$$P(z_t \in A | x_t = x) = \int_A p_t(z|x;\theta) dz, \quad t = 1, \dots, T, \quad (14)$$

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

Let us now for intuitiveness consider the case of the heuristic switching model suggested in Section 3. For the given NKM specification, $z_t = \{y_t, \pi_t, r_t\}$; therefore, $l = 3$ following equations (1) to (3). x_t only contains lagged dependent variables z_t , i.e., no other exogenous explanatory variables are considered in this model specification. ε_t stands for a set of l idiosyncratic shocks $\{\varepsilon_{y,t}, \varepsilon_{\pi,t}, \varepsilon_{r,t}\}$ that are i.i.d. around mean zero with variance $\sigma_{\{y,\pi,r\}}^2$. Finally, the set of estimated parameters contains all *structural* and *bounded rationality* coefficients, i.e., $\theta = \{\tau, \kappa, \phi_y, \phi_\pi, \eta, \iota, \mu, \gamma\}$ and $n = 8$. Recall that for the intrinsic persistence parameters $\chi = \alpha = \phi_r = 0$ hold, and the remaining model

parameters are parametrized. In contrast, for the hybrid RE NKM, $\theta = \{\chi, \alpha, \tau, \kappa, \phi_y, \phi_\pi, \phi_r\}$, i.e. $n = 7$.

As a result of the adaptive revisions of the expectations introduced by equation (9), the probability $p_t(z|x;\theta)$ in equation (14) does not have a closed-form representation. Therefore, an exact mathematical derivation of the likelihood function of the model in equation (13) does not exist, and a standard estimator of θ , the maximizer of the conditional log-likelihoods

$$\tilde{\theta} = \arg \underbrace{\max}_{\theta \in \Theta} L_T(\theta), \quad (15)$$

where $L_T(\theta) = \sum_{t=1}^T \log p_t(z_t|x_t;\theta)$, is infeasible.³ However, we are always able to obtain simulated observations from the model (13). The SML method presented below then allows us to numerically 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 $p_t(z_t|x_t;\theta)$, $t = 1, \dots, T$, we first generate $N \times T$, $N \in \mathbb{N}$, i.i.d. draws from the l -dimensional distribution \mathcal{F}_ε , and $\{\varepsilon_{i,t}\}_{i=1}^N$ to compute

$$Z_{i,t}^\theta = m_t(x_t, \varepsilon_{i,t}, \theta), \quad i = 1, \dots, N. \quad (16)$$

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

$$\hat{p}_t(z_t|x_t;\theta) = \frac{1}{N} \sum_{i=1}^N K_H(Z_{i,t}^\theta - z_t), \quad (17)$$

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 $l \times l$ bandwidth matrix.

Using the simulated conditional density $\hat{p}_t(z_t|x_t;\theta)$, we can derive the SML estimator of θ :

$$\hat{\theta} = \arg \underbrace{\max}_{\theta \in \Theta} \hat{L}_T(\theta), \quad (18)$$

³Moreover, 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.

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 $\mathcal{F}_\varepsilon(\cdot)$, $\{\varepsilon_{i,t}\}_{i=1}^N$ for all values of θ . If $\hat{L}_T(\theta)$ is continuous and differentiable in θ , then numerical optimization is facilitated. Considering equation (17), if K and $\theta \mapsto m_t(x_z, \varepsilon_t, \theta)$ are $q \geq 0$ continuously differentiable, then the same holds for $\hat{L}_T(\theta)$. Under the regularity conditions on the conditional density p_t and kernel K (Kristensen and Shin, 2012, conditions A.1-4, K.1-2, pg. 80–81), $\hat{p}_t(z_t|x_t; \theta) \xrightarrow{P} p_t(z_t|x_t; \theta)$, which implies that $\hat{L}_T(\theta) \xrightarrow{P} L_T(\theta)$ as $N \rightarrow \infty$ for a given $T \geq 1$. Thus, the SML estimator, $\hat{\theta}$, retains the same properties as the infeasible MLE, $\tilde{\theta}$, as $T, N \rightarrow \infty$ under suitable conditions. Other important properties of the SML estimator are discussed in the [Online Appendix A](#).

5. Monte Carlo Study

This section numerically investigates the finite-sample properties of the SML estimator in a macroeconomic framework. We conduct an extensive Monte Carlo simulation study to determine its capability to consistently recover the pseudo-true parameters in a controlled environment. We focus on issues such as the sources of estimation bias and uncertainty, on potential specifications and parametrizations of the model, and on the computational setup of the optimization procedure. The following two sections summarize the general settings for all numerical exercises if not explicitly stated otherwise.⁴

5.1. Model Parametrization for the Study

Starting with the BR NKM, we first parametrize the discount factor $\nu = 0.99$ whose value has strong support in the related empirical estimation literature. Other structural parameters (τ , κ , ϕ_y , ϕ_π), which are subject to estimation, and the standard deviations of the idiosyncratic shocks ($\sigma_y = 0.543$, $\sigma_\pi = 0.240$, $\sigma_r = 0.151$) are parametrized according to the recent results for the US data by [Jang and Sacht \(2019, Table 2, EFB scenario\)](#). The idiosyncratic shocks $\varepsilon_{\{y, \pi, r\}}$ are

⁴All computations are conducted using Julia version 1.4.0 (2020-03-21). The computational burden becomes manageable by utilizing the [Distributed.jl](#) package for parallel computing on multi-core computers/servers. For optimizations, the [BFGS](#) algorithm from the [Optim.jl](#) package is used with an additional [LineSearches.jl](#) ([.BackTracking](#)) functionality to decide the step length. BFGS is a gradient-based iterative optimization algorithm that searches for a local minimum of a differentiable function. In our specification, it minimizes the negative simulated log-likelihood function $\hat{L}_T(\theta)$. BFGS approximates the Hessian using differences in the gradient across iterations. The starting matrix is an identity matrix. The BFGS algorithm was tested against other algorithms available in the [Optim.jl](#) package, and it delivers the best performance for the given optimization problem. It also delivers a largely comparable performance to differential evolution algorithms from an alternative [BlackBoxOptim.jl](#) package.

therefore independently drawn from a normal distribution with mean zero and variance $\sigma_{\{y,\pi,r\}}^2$, which is a standard and reasonably realistic assumption for the distribution \mathcal{F}_ε in equation (13). We also take advantage of the favorable theoretical properties of the Gaussian kernel (Kristensen and Shin, 2012, pg. 81) related to equation (17). The standard deviations of the idiosyncratic shocks $\sigma_{\{y,\pi,r\}}$ cannot be subject to estimation as the SML method requires the distribution \mathcal{F}_ε of the i.i.d. shocks to be known.

The BR parameters (η, ι, μ) follow the parametrization by Hommes et al. (2019) except that our trend-following parameter ι is defined as an average of their “weak” and “strong” trend-following parameters to avoid issues with parameter identification. The parameter for the intensity of choice is set to $\gamma = 1$, which follows an economic intuition of a low intensity of the switching of agents between forecasting heuristics.⁵ Finally, no memory of the past forecast performance is assumed, $\rho = 0$, based on the results of Jang and Sacht (2016, 2019), who generally find this parameter to be insignificant for the heuristic switching model using the SMM estimation technique. However, as other purely simulation-based studies (De Grauwe, 2011; Hommes et al., 2019; De Grauwe and Ji, 2020) generally calibrate the memory parameter to non-zero values, an additional sensitivity analysis for $\rho = \{0.3, 0.5, 0.7\}$ is provided in Subsection 5.3.5.

For the hybrid RE NKM, we again calibrate the discount factor by $\nu = 0.99$. The structural parameters together with the standard deviations of the shocks follow common parametrizations based on Hommes et al. (2019) and De Grauwe and Ji (2020).

Accordingly, the values of the pseudo-true parameters that are subject to estimation are listed in Table 1. However, the qualitative results presented are robust with respect to other realistic parametrizations of the model, e.g., based on additional results by Jang and Sacht (2019) and parametrization setups by Hommes et al. (2019) and De Grauwe and Ji (2020) to which we devote attention in Subsection 5.3.5.

⁵Our choice of the numerical value of γ closely follows the related literature. For comparison, De Grauwe (2010, 2011, 2012); Liu and Minford (2014); Jang and Sacht (2019) also calibrate $\gamma = 1$, Hommes et al. (2019) calibrate this parameter to 0.4, De Grauwe and Ji (2020) set $\gamma = 2$, while Grazzini et al. (2017) use $\gamma = 5$ when estimating the De Grauwe (2012) model. Jang and Sacht (2016) employ a range $\gamma = \{0.1, 1, 10\}$ for negligible, smooth and slow, and strong switching, respectively, to estimate the De Grauwe (2011) model. We recall that a negative γ lacks economic sense, while a high value of this parameter would imply an unrealistic speed of opinion updating that is unlikely in macroeconomic reality for the quarterly periodicity of data. Indeed, Jang and Sacht (2016) use $\gamma = 100$ as an approximation to infinity. Importantly, a high γ value may also lead to a numerical divergence of the model.

5.2. Simulation Setup

We study the performance of the SML estimator under three lengths of time series generated from the model, namely, 250, 500, and 5,000. This allows us to maintain an admissible time span for common quarterly macroeconomic data while studying the asymptotic tendencies of the estimator with an increasing sample size. Moreover, 1,000 additional initial simulated observations are always discarded as a burn-in period. The random seed is controlled at the level of individual Monte Carlo runs, which allows for the replicability of the results.

In the general setup, we apply a constrained optimization of a multivariable function and jointly estimate 8 parameters for the BR NKM or 7 parameters for the hybrid RE NKM. The parameter space is restricted to the intervals summarized in [Table 1](#). The constraints are based on the theoretical borders for given parameters (cf. [Section 3](#)) combined with a preliminary rough search based on a broader space to verify the sufficient lengths of the intervals. The random starting point for the optimization search is uniformly generated from given intervals. The precision of the kernel density approximation is set to $N = 1000$, and to ensure the statistical validity of the results while keeping the computational burden manageable, 300 independent runs are always conducted. 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 $l = 3$, $s = \{1, \dots, l\}$, $\hat{\sigma}_s$ denotes the sample standard deviation of the elements of the s^{th} dimension of $\{\varepsilon_i\}_{i=1}^N$ and off-diagonal terms $H_{s_1,s_2} = 0$, $s_1 \neq s_2$.

5.3. Numerical Results

We primarily investigate the capability of the SML estimator to recover the pseudo-true parameters in finite samples. The focus is on the estimation precision for the BR NKM and on a discussion of the probable sources of a potential estimation bias. Another dimension of the analysis is a numerical evaluation of the asymptotic tendencies to the consistency and efficiency of the estimator. A comparison with the hybrid RE NKM is also provided.

5.3.1. Estimation Accuracy

[Table 1](#) reveals a very promising performance of the SML estimator. An alternative graphical depiction of the identical results is provided in [Figure 1](#). First, by focusing on the BR NKM in the

Table 1: Results of the Monte Carlo simulations

| | Par. | BR NKM | | | Par. | Hybrid RE NKM | | |
|------------------------|-------|-------------|-------------|-------------|------|---------------|-------------|-------------|
| | | $T=250$ | 500 | 5000 | | $T=250$ | 500 | 5000 |
| χ | — | — | — | — | .50 | .53 | .53 | .53 |
| $\langle 0, 1 \rangle$ | | | | | | (.31-.83) | (.37-.70) | (.48-.59) |
| α | — | — | — | — | .50 | .53 | .52 | .53 |
| $\langle 0, 1 \rangle$ | | | | | | (.27-.85) | (.37-.73) | (.48-.60) |
| τ | .371 | .48 | .47 | .47 | .20 | .23 | .22 | .22 |
| $\langle 0, 1 \rangle$ | | (.38-.58) | (.39-.55) | (.44-.49) | | (.13-.35) | (.15-.31) | (.20-.24) |
| κ | .213 | .23 | .23 | .23 | .30 | .36 | .35 | .35 |
| $\langle 0, 1 \rangle$ | | (.16-.30) | (.18-.28) | (.22-.24) | | (.22-.51) | (.26-.46) | (.32-.38) |
| ϕ_y | .709 | .71 | .71 | .71 | .50 | .49 | .47 | .47 |
| $\langle 0, 1 \rangle$ | | (.69-.74) | (.69-.73) | (.70-.72) | | (.18-.85) | (.25-.70) | (.39-.54) |
| ϕ_π | 1.914 | 1.91 | 1.91 | 1.91 | 1.50 | 1.62 | 1.61 | 1.62 |
| $\langle 1, 3 \rangle$ | | (1.87-1.94) | (1.89-1.93) | (1.90-1.92) | | (1.20-2.06) | (1.37-1.92) | (1.54-1.70) |
| ϕ_r | — | — | — | — | .50 | .49 | .49 | .49 |
| $\langle 0, 1 \rangle$ | | | | | | (.37-.58) | (.42-.56) | (.47-.51) |
| η | .65 | .67 | .66 | .67 | — | — | — | — |
| $\langle 0, 1 \rangle$ | | (.28-1) | (.51-.90) | (.62-.71) | | | | |
| ι | .85 | .90 | .88 | .87 | — | — | — | — |
| $\langle 0, 2 \rangle$ | | (.39-1.26) | (.61-1.13) | (.81-.93) | | | | |
| μ | .50 | .53 | .54 | .56 | — | — | — | — |
| $\langle 0, 1 \rangle$ | | (.34-.76) | (.42-.67) | (.52-.60) | | | | |
| γ | 1.00 | .98 | 1.04 | 1.01 | — | — | — | — |
| $\langle 0, 5 \rangle$ | | (.08-2.98) | (.40-1.98) | (.83-1.21) | | | | |

The constraints for optimization and its starting point are given in $\langle \rangle$ brackets. T denotes the length of the executed time series. The sample medians based on 300 random runs are reported while the 95% confidence intervals of the sample estimates are reported in $()$ parentheses. The parametrization is based on [Jang and Sacht \(2019\)](#), Table 2, EFB, for the BR NKM and on [Hommes et al. \(2019\)](#); [De Grauwe and Ji \(2020\)](#) for the hybrid RE NKM. The figures are rounded to 2 or 3 decimal places.

left half of the table/figure, the values of the Taylor rule coefficients ϕ_y and ϕ_π are recovered nearly perfectly, and the 95% sample confidence intervals suggest a minimal variance of the estimator for the given parameters. The slope of the New-Keynesian Phillips Curve κ is also estimated very precisely. For the elasticity of substitution in consumption τ , we observe positively biased results that suggest a stable bias of the SML estimator for this parameter. The sampling distributions have regular symmetric shapes with a skewness close to zero and kurtosis close to a normal distribution.

The bias for τ can be explained at the level of the model structure as follows. Equation (1) essentially plugs r_t (3) into the dynamic IS curve where it directly interacts with τ . This creates a composite error term $\varepsilon_{y,t} + \tau\varepsilon_{r,t}$ that introduces a correlation between the individual errors in the model. It also naturally correlates with τ . Moreover, τ thus also loads directly on the Taylor rule coefficients and indirectly on κ through the $\phi_y y_t$ term in equation (3). All

of this leads to a problematic identification of τ that results in a bias of the estimator for this parameter. This seems to be a fundamental finding gained for the interpretation of the empirical results, where unfortunately, we can hardly predict even the direction of the bias for τ . One potential solution analyzed in [Subsection 5.3.5](#) is to calibrate this parameter. It is important to emphasize, however, that in such a complicated nonlinear system as the BR NKM, similar biases are practically inevitable. The purpose of Monte Carlo situations is exactly to learn about their existence as concretely as possible through numerical simulations.

[Place [Figure 1](#) around here.]

The BR parameters are generally more challenging to estimate, but the SML method still delivers a very good performance. The sample variances for η , ι , and μ are relatively larger than for the structural parameters; however, the sampling distributions have regular and reasonably symmetric shapes with kurtosis generally higher compared to a normal distribution. We also do not observe any considerable bias tendencies except for a small positive bias for μ . The most important result is undoubtedly the estimation performance for the switching parameter of the intensity of choice γ . Even in simple univariate financial models, capturing the effect of this coefficient is generally difficult ([Boswijk et al., 2007](#); [Hommes, 2013](#); [Lamperti, 2018a](#); [Lux and Zwickels, 2018](#)). In addition, a handful of previous macroeconomic studies have had difficulties estimating this parameter (cf. [Subsection 2.2](#)). On the contrary, our results show that the SML estimator is well capable of estimating the pseudo-true intensity of choice γ under the multinomial logistic updating scheme (9). That is, compared to other estimation methods applied to date, the SML estimator demonstrates a considerable ability to estimate γ without any bias and without issues with statistical insignificance at the standard 5% level. It is thus capable of detecting signs of behavioral switching in the simulated model output.

5.3.2. *Asymptotic Tendencies*

Favorable asymptotic tendencies of the estimator are apparent with an increasing sample size $T = 250, 500, 5000$. We generally observe a considerable narrowing of the 95% confidence intervals of the sample estimates when T goes from 250 to 500 and further to 5,000. This tendency leads to

markedly stronger improvements of the estimation efficiency for the BR parameters compared to the structural parameters. The results further demonstrate significant improvements in the estimation precision for the intensity of choice γ with an increasing sample size. Moreover, although we observe a positive skew of the sampling distribution of the estimator for γ for $T = 250$, this asymmetry decreases for $T = 500$ and disappears completely for $T = 5000$. The same normalization of the shape of the sampling distribution can be observed at the level of kurtosis. For $T = 250, 500$, the excess kurtosis is slightly positive for the structural parameters and rather large for the BR parameters, whereas for $T = 5000$, such a non-normal shape disappears with the excess kurtosis for all parameters approaching zero. In contrast, the bias of the SML estimator for parameters τ and μ appears to be stable regardless of the sample size. A very subtle tendency towards bias reduction can be observed for a small bias for ι .

5.3.3. Shape of the Log-Likelihood Function

A set of regularity conditions A.1-A.4, that impose restrictions on the model and the conditional density, are defined in [Kristensen and Shin \(2012\)](#) such that the estimated conditional density \hat{p} converges sufficiently fast to the true conditional density p . Therefore, the asymptotic equivalence of the estimated $\hat{\theta}$ and the true θ parameter vectors are assured. [Kristensen and Shin \(2012\)](#) assert that these assumptions are “quite weak and are satisfied by many models”; nonetheless, the analytical intractability of the analyzed heuristic switching macroeconomic model does not allow us to mathematically verify these conditions. Therefore, we take advantage of 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.

[Place [Figure 2](#) around here.]

The left half of [Figure 2](#) shows the shape of the simulated log-likelihood function for the BR NKM. As we are unable to graphically depict an 8-dimensional object, the individual curves show the transversal profiles of the simulated log-likelihood function in the planes of given parameters. Other parameters are held fixed at their parametrized values ([Table 1](#)). A smooth shape of the surface that results from the use of a Gaussian kernel based on $N = 1000$ approximation points

is clearly observable for all structural parameters and the intensity of choice γ over the entire parameter space. Moreover, the unique maxima are well detectable for the structural parameters with the strongest curvature in the direction of the parameter ϕ_π , which naturally corresponds to the numerical results reported in [Table 1](#) and [Figure 1](#). For the remaining BR parameters except for γ , one can observe considerable differences and irregularities between some of the independent realizations of the simulated log-likelihood function. Together with a rather flat surface of the average likelihood function, these differences are likely the cause of higher sample variances for η , ι , and μ . This observation explains our findings regarding the more challenging estimation of these parameters. The bias for τ is also obvious from the shape of the simulated log-likelihood. When increasing the sample size from $T = 500$ (panel c) to $T = 5000$ (panel e), we obtain distinctively sharper shapes and also a more regular behavior of the log-likelihood for all independent runs. Accordingly, based on the smooth surfaces of the average simulated log-likelihood functions and unique maxima observed in the directions of all estimated parameters, we can generally assume that the regularity conditions and identification of the parameters are satisfied for the BR NKM.

5.3.4. Comparison with the RE Model

The results for the hybrid RE NKM are also summarized in [Table 1](#) and the right half of [Figure 1](#). The estimation performance is generally worse for the hybrid RE NKM than for the BR NKM, while the number of estimated parameters is one smaller. This result might appear counterintuitive at first, since the BR NKM is expected to generate more complex dynamics than the RE NKM. One reason for this finding might be the qualitatively different behavior of the output time series of the realizations in the RE case. In contrast, in the model under BR, the output displays a considerable stability and path dependency that emerges from the incorporation of behavioral heuristics, and the fluctuations of the output time series of the hybrid RE NKM are more intense, regardless of the impact of the intrinsic persistence parameters χ , α and ϕ_r . Exogenous random shocks rather than (cross-)autocorrelation structures thus seem to be a driver for the hybrid RE NKM dynamics, which would naturally complicate the proper detection of the pseudo-true parameter values for any estimation method. The inability of the SML estimator to approximate the likelihood function accurately enough for more volatile time series is widely

studied in [Kukacka and Barunik \(2017\)](#). In such cases, the SML criterion function displays a flatter surface that hinders the optimization algorithm from reaching the global optimum in a multidimensional parameter space. This can generally be observed in the right half of [Figure 2](#), especially for parameters κ , ϕ_y , and ϕ_π in which the individual courses of the depicted shapes are more dispersed.

Another related issue regards small biases of the estimator that are mostly visible for parameters κ and ϕ_π . Moreover, the bias does not tend to disappear as the sample size increases, which suggests unfavorable asymptotic tendencies. In terms of efficiency, the 95% sample estimate intervals are rather large for the intrinsic persistence parameters χ and α and also for the Taylor rule coefficients ϕ_π and ϕ_y . However, for ϕ_π and ϕ_y the estimation performance is incomparable to that of the BR NKM as the interest rate smoothing parameter ϕ_r takes over their impact. The relative improvement in efficiency as the sample size increases seems comparable.

5.3.5. Additional Model Specifications

To support the qualitative robustness of our findings, we repeat the same numerical analysis for additional alternative realistic parametrizations for both the BR and the hybrid RE NKM. Because of space constraints, in Appendix A, we only report the outputs based on the $T = 500$ sample size, which is comparable to [Figure 1](#), panels (c) and (d). First, [Figure 3](#) reports the results for the BR NKM parametrizations based on the estimates for the Euro Area data by [Jang and Sacht \(2019, Table 3, EFB scenario\)](#) and based on the parametrization setups of [Hommes et al. \(2019\)](#); [De Grauwe and Ji \(2020\)](#) with the value of the parameter $\tau = 1$. In both cases, we observe an estimation performance largely comparable to our benchmark setup ([Figure 1](#)). The bias for τ is also apparent for both parametrizations. Interestingly, for the [Jang and Sacht \(2019, Table 3, EFB scenario\)](#) parametrization, the bias for μ disappears, while for the [Hommes et al. \(2019\)](#); [De Grauwe and Ji \(2020\)](#) parametrization, we instead detect new biases for the Taylor rule parameters ϕ_y and ϕ_π . The same similarity of the qualitative results holds for the alternative parametrizations of the hybrid RE NKM based on [Jang and Sacht \(2019, Tables 2 and 3, REH scenario\)](#); [De Grauwe and Ji \(2020\)](#) reported in [Figure 4](#). Interestingly, under the given parametrization, the SML estimator seems to be more accurate for α , and also the small biases for the structural parameters apparent

in [Figure 1](#) seem to diminish. However, τ generally suffers from statistical insignificance, which is due to its very small pseudo-true values while maintaining the same estimation interval $\langle 0, 1 \rangle$ for comparability. We can thus conclude that the properties of the SML estimator are not only model-dependent but also parametrization-dependent.

Second, to address the issue of the consistent bias for the inverse intertemporal elasticity of substitution in consumption τ for the BR NKM, we parametrize its value, and only 7 remaining parameters are now subject to estimation. The results for the three alternative parametrizations of τ are depicted in [Figure 5](#). Although holding this weekly identified coefficient constant does not lead to a better estimation performance for the other parameters compared to the benchmark case, the bias does not spill over to the other parameters, and the issue therefore seems to be well resolved.

Third, we also experiment with parametrizing the memory parameter ρ in equation (8) to nonzero values, specifically to $\rho = \{0.3, 0.5, 0.7\}$. We recall that [Jang and Sacht \(2016\)](#) find this memory coefficient to be statistically insignificant from estimation via the SMM. [Grazzini et al. \(2017\)](#) also report $\rho = 0$ to have a negligible effect for the model behavior as its posterior distribution does not depart significantly from the prior distribution in a Bayesian estimation. Setting ρ to nonzero implies the issue of the latent dynamics in the model in which the conditional log-likelihood $\hat{L}_T(\theta)$ is infeasible. However, here, we are instead interested in what numerical issues this theoretical imperfection brings. The “experimental” results reported in the [Online Appendix B](#), [Figure 7](#), clearly show that increasing ρ first deteriorates the estimation performance for all BR parameters ($\rho = 0.3$), it further deteriorates the estimation of the structural coefficients τ and κ , it makes the estimates of the BR parameters unreliable ($\rho = 0.5$), and it finally leads to a complete loss in the estimation performance of the SML estimator for all coefficients except the Taylor rule coefficients ($\rho = 0.7$). The explanation is twofold. In addition to the latent dynamics issue, the memory parameter ρ in equation (8) interacts with the intensity of choice parameter γ in equation (9) through $U_{x,t}^k$. If both parameters are estimated simultaneously, then a serious joint-identification problem arises. In our case, however, the memory parameter ρ is parametrized. The interaction of the two parameters thus leads only to a suppression of the effect of the intensity

of choice γ on the model dynamics. This, in turn, causes the abovementioned effects.

Finally, in Appendix B, [Figure 6](#), we report additional results for the application of the SML estimator to different heuristic switching models derived based on alternative sets of heuristics. Specifically, we estimate the models following the heuristic structure by [Gaunersdorfer and Hommes \(2007\)](#); [De Grauwe \(2011\)](#); [De Grauwe and Ji \(2020\)](#) and with τ parametrized to avoid the estimation bias. The accuracy and efficiency of the SML estimator for these alternative models is largely comparable to our findings in the previous sections. This provides an additional strong robustness check to our Monte Carlo simulation study.⁶

6. Empirical Applications

This section examines the empirical performance of the SML in a multivariate macroeconomic setup. We are also interested in evaluating and reporting the parameter estimates used for activities in policy advisory. Both the BR NKM and its hybrid RE NKM counterpart are considered since a purely forward-looking RE NKM without lag terms does not capture the empirically observed inertia in inflation and output due to the aforementioned “persistence anomaly” ([Chari et al., 2002](#)). Since the data for the US economy exhibit a high degree of persistence, it is worthwhile to investigate different model specifications where inertia is technically incorporated either by heuristics (BR NKM) or by the assumption regarding habit formation and price indexation (hybrid RE NKM).

6.1. Estimation Setup

The algorithm for empirical estimation generally adopts the setup of the simulation study in subsections [5.1](#) and [5.2](#) with two minor alterations. First, 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 extend the interval for the Taylor rule coefficient ϕ_π to $\langle 0, 3 \rangle$, which supports an optimization search. Second, we increase the precision of the kernel density approximation to $N = 2000$. For comparison exercises, we consider three additional parameter sets where we parametrize the structural parameters τ and/or κ as summarized in [Table 2](#).

⁶As an additional robustness exercise, [Online Appendix C](#), Table 4 and Figure 8, report the results designed for a direct numerical comparison of the estimation performance between the SML method and the SMM, which suggests the superiority of the SML estimator at least for the given model setup.

Table 2: Results of the empirical study for US data: BR vs. hybrid RE NKM

| | BR NKM | | | Hybrid RE NKM | | |
|-------------------------|-------------|-------------|-------------|---------------|------------|-------------|
| | All | B | C | All | D | All via MLE |
| χ | — | — | — | .99 | .74 | 1.00 |
| $\langle 0, 1 \rangle$ | | | | (.81-1.00) | (.67-.999) | (.81-1.00) |
| α | — | — | — | .77 | .999 | .81 |
| $\langle 0, 1 \rangle$ | | | | (.70-.83) | (.90-1.00) | (.76-.86) |
| τ | .00 | <i>.371</i> | <i>.20</i> | .05 | .02 | .04 |
| $\langle 0, 1 \rangle$ | (.00-.00) | | | (.02-.08) | (.004-.04) | (.03-.05) |
| κ | .09 | <i>.213</i> | <i>.05</i> | .00 | <i>.03</i> | .00 |
| $\langle 0, 1 \rangle$ | (.007-.17) | | | (.00-.00) | | (.00-.003) |
| ϕ_y | .12 | .05 | .04 | .78 | .69 | .79 |
| $\langle 0, 1 \rangle$ | (.10-.14) | (.03-.07) | (.01-.06) | (.60-.98) | (.45-.93) | (.50-1.08) |
| ϕ_π | 1.26 | 1.23 | 1.27 | 1.28 | 1.32 | 1.51 |
| $\langle 0, 3 \rangle$ | (1.14-1.31) | (1.19-1.29) | (1.22-1.33) | (1.03-1.54) | (.26-1.58) | (1.26-1.76) |
| ϕ_r | — | — | — | .91 | .90 | .93 |
| $\langle 0, 1 \rangle$ | | | | (.88-.93) | (.87-.93) | (.91-.95) |
| η | .19 | .21 | .20 | — | — | — |
| $\langle 0, 1 \rangle$ | (.04-.27) | (.09-.25) | (.09-.26) | | | |
| ι | .00 | .00 | .00 | — | — | — |
| $\langle 0, 1 \rangle$ | (.00-.002) | (.00-.02) | (.00-.05) | | | |
| μ | .31 | .38 | .44 | — | — | — |
| $\langle 0, 1 \rangle$ | (.07-.67) | (.15-.62) | (.19-.65) | | | |
| γ | 1.40 | 1.49 | 1.39 | — | — | — |
| $\langle 0, 10 \rangle$ | (.50-7.26) | (.79-6.22) | (.78-6.35) | | | |

The constraints for optimization and its starting point are given in $\langle \rangle$ brackets. The sample medians based on 300 random runs are reported, while the 95% confidence intervals of the sample estimates are reported in $()$ parentheses. The parametrization is based on [Jang and Sacht \(2019\)](#), Table 2, EFB (for B), the hybrid RE (for D), and [De Grauwe and Ji \(2020\)](#) for C. The fixed parameters are marked in italics. The figures are rounded to 2 or 3 decimal places.

6.2. Data

We obtain the US quarterly data from the website of the Federal Reserve Bank of St. Louis.⁷ Inflation is measured using the seasonally adjusted consumer price index (for all urban consumers & items). Output is obtained from seasonally adjusted real GDP based on billions of chained 2012 dollars. A standard smoothing parameter of $\lambda = 1600$ is used to estimate the trend of the observed data from the one-sided Hodrick-Prescott filter ([Stock and Watson, 1999](#)) for output. The effective federal funds rate is used to measure the short-term nominal interest rate in the US. The sample covers the period from 1954:Q3 to 2019:Q2 with 260 observations in total.

6.3. Empirical Results

The parameter estimates for both model specifications are shown in [Table 2](#). We first focus on the BR NKM for which all BR parameters are statistically significant (except for the trend-following parameter ι ; see below). Note that the column entitled “All” reports the results for which none of the parameters are parametrized to a pre-defined fixed value. Under this scenario, the adaptation parameter η exhibits a median value of 0.19, i.e., approximately 20% of the past realization in the output gap, and the inflation rate is considered in case the adaptive heuristic (4) is applied. Therefore, a naïve expectation formation process, with η being close or equal to unity, can be ruled out. The estimate of the extrapolation parameter μ given by 0.31 reveals a weak importance of the deviations in the output gap and of the inflation rate from the corresponding anchor point, which is defined as a sample average of the previous realizations. This median value of μ from the heuristic (6) comes close to the one proposed by [Hommes et al. \(2019\)](#).

A novelty of our empirical investigation is that we are reliably able to identify the intensity of choice parameter γ in a macroeconomic model setup. We obtain a clearly statistically significant estimate of 1.40. This implies that the agents have the incentive to revise their expectation formation process in every period. According to [De Grauwe and Ji \(2020\)](#), γ can be interpreted as “expressing a willingness to learn from past performance”. Based on our estimate, this degree of willingness turns out to be rather low. As a result, the switching from one forecasting heuristic to the other becomes smooth and slow considering the quarterly frequency of the data.⁸

We now briefly discuss the estimate for ι , which denotes the extrapolation parameter in the trend-following heuristic (5). It measures how strongly the expectation formation based on this specific forecasting rule considers the change in the direction of the output gap or the inflation rate up to the second lag. The observation that ι turns out to be insignificant comes not as a surprise. The reason for this is twofold. First, technically speaking, one can note that the corresponding term $(x_{t-1} - x_{t-2})$ can be found in both heuristics for trend-following and learning anchoring and

⁷ Available at fred.stlouisfed.org [accessed 9 September 2019].

⁸ Note that the results regarding the intensity of choice γ can be compared to those presented for similar representations of the BR NKM obtained via the SMM in [Jang and Sacht \(2016, 2019\)](#), where a Taylor rule with interest rate smoothing is assumed. The intensity of choice parameter is fixed in their study, which means that as a cross-check, the case of $\gamma = \{0.1, 1\}$ for the US should be considered since we find the median value of γ close to unity.

adjustment, i.e., (5) and (6), respectively. The difference lies in the weighting scheme of ι and unity. An insignificant estimate for ι therefore indicates the redundancy of an additional but identical term as part of both forecasting rules. Second, note that the following economic interpretation holds with respect to heuristic (5) in isolation. By acknowledging the extrapolation effect, it gives rise to strong oscillatory dynamics for a high value of ι (Anufriev and Hommes, 2012). This resembles a destabilized economy with high degrees of fluctuation in the macroeconomic variables. However, the trend-following heuristic hardly survives if the central bank reacts strongly to disturbances in output and inflation (Hommes, 2021). This is, in fact, suggested by our estimates for ϕ_y and ϕ_π . Instead, the agents' expectation formation is dominated by the other two heuristics, namely, *ADA* (4) and *LAA* (6). Our empirical results thus support the experimental findings by (Hommes et al., 2019, Figure 11) who also find that these two heuristics strongly dominate various experimental setups. According to Assenza et al. (2013), as the central bank successfully does its job in stabilizing the economy over time, changes in $(x_{t-1} - x_{t-2})$ flatten out. Such negative-feedback policies of the central bank prevent coordination on the trend-extrapolating behavior and the survival of trend-following strategies (Assenza et al., 2019). This makes the term in the *TR* heuristic (5) most likely obsolete when it comes to expectation formation, which is then reflected by ι not deviating significantly from zero.

A discussion on the estimates for the structural parameters follows. We focus again on the “All” scenario first. Although the estimates of the monetary policy parameters indicate a strong emphasis on inflation over output stabilization with median values of $\phi_y = 0.12$ and $\phi_\pi = 1.26$, the remaining structural parameters τ and κ turn out to be (close to) statistically insignificant. This observation regarding the slopes of the dynamic IS equation (1) and the New-Keynesian Phillips Curve (2) holds under the inspection of the corresponding 95 % confidence intervals. At least for τ , the result ties into our discussion in Subsection 5.3.1 concerning a weak identification and the estimation bias for this parameter. It follows that the demand and supply curves seem to be decomposed from the remaining system of equations and that inherited persistence, in terms of cross-volatility within the inflation-output gap nexus, is largely absent.

As output and inflation dynamics seem to be only characterized by the parametrized shocks

and heuristics, this raises the issue of a potential model mis-specification. Although the BR NKM is by definition a perfectly specified model in the Monte Carlo environment, it struggles when dealing with an empirical data set that exhibits structural breaks in inflation volatility. Such structural breaks are, however, a widely known phenomenon for the US economy in the second half of the 20th century.

The following exercise reveals that the estimates of all additional parameters turn out to be robust with respect to different parametrizations of τ and κ . Therefore, we report two additional sets of estimates denoted by “B” and “C” in the third and fourth columns of [Table 2](#), respectively. For “B”, both critical parameters are parametrized according to [Jang and Sacht \(2019, Table 2, EFB scenario\)](#). For “C”, we apply a plain vanilla parametrization following [De Grauwe \(2011, p. 452\)](#). We choose both parametrizations since these are directly linked to the empirically examined BR NKMs similar to the one considered in this paper.

6.3.1. Hybrid RE NKM

We now shed light on the empirical results linked to the hybrid RE NKM. We consider two different sets. In the set “All”, none of the parameters are parametrized. For “D”, we parametrize κ , which happens to be statistically insignificant under the “All” set. We choose a standard value of 0.03 taken from the literature on RE NKMs in general. For κ being parametrized, τ now becomes nearly insignificant as seen in the “D” column of [Table 2](#). This is, once again, an indication of the distorted cross-correlation between output and inflation. Potential explanations include not only a mis-specification of the theoretical model but also a weak identification of κ in general. This might be caused by the existence of structural breaks in the underlying time series for the US as discussed above. However, this seems to be a puzzling observation as, like in the BR NKM case, the estimation method performs well in simulations. We conclude that a further exploration of this issue is needed in future research.

We now offer a brief economic interpretation of the estimates for the hybrid RE NKM. The median values for χ , α , and ϕ_r under the set “All” come close to their upper bounds, which are given by unity in all cases. This reflects a high degree of persistence in the underlying empirical time series. It must then be mimicked by a high degree of hybridity in the theoretical model,

i.e., high values of the parameters for habit formation (χ), price indexation (α), and interest rate smoothing (ϕ_r). The result of $\phi_\pi = 1.28$ reveals a value that is virtually identical to the one obtained in the BR NKM estimates. With respect to ϕ_y , it attains a comparatively high value of 0.78, which also implies a strong focus on output stabilization.

For completeness, we take advantage of the linearity of the hybrid RE NKM and estimate the model via the maximum likelihood technique. The classical MLE based on the true likelihood is derived in Appendix C. We consider the “All via MLE” set where no parameters are parametrized. The results are reported in the last column of Table 2. In a direct comparison to the estimates obtained via the SML method, these are hardly distinguishable from the estimates obtained through the MLE. The only exception are the estimates for ϕ_π . In addition, for the MLE the width of the 95% confidence interval is narrower compared to the “D” column but is similar to the “All” column. Overall, it can be concluded that the SML approach approximates the MLE method very well.

7. Conclusion

In the absence of the RE paradigm, a growing number of studies on DSGE models address the importance of boundedly rational expectation formation. This paper examines a baseline NKM with heterogeneous agents who adopt simple heuristics in forecasting future movements of output and inflation. In this heuristic switching model by Hommes et al. (2019), different groups of agents form expectations according to an adaptive, a trend-following, or a learning anchoring and adjustment scheme. The corresponding nonlinear specification of the model is for the first time estimated via the SML method (Kristensen and Shin, 2012). This approach relaxes some restrictive theoretical assumptions required by competing estimation methods and further reduces the discretionary choices necessary for their practical implementation. We modify the univariate version of the SML applied in financial econometrics and transfer it to multivariate macroeconomic optimization. The results are compared to the findings obtained for a hybrid NKM with a lead and lag structure under RE.

In a Monte Carlo simulation study, we investigate the finite-sample properties of the SML estimator in a multivariate framework. An important advantage lies in its ability to estimate

the intensity of choice that governs the multinomial logistic switching process from one group to another. In previous related empirical studies, this parameter had to be parametrized or was estimated as potentially statistically insignificant or uninformative for the model behavior. On the contrary, our results show that the SML can estimate the switching parameter with surprising precision and with no small sample bias. Moreover, all three additional behavioral parameters and most of the structural parameters are also estimated very precisely. Accordingly, our numerical analysis confirms a strong capability of the SML method to estimate the heuristic switching model parameters.

In an empirical application, we estimate the model in its BR version and its hybrid counterpart under RE. Based on the US economy’s data covering the period from 1954:Q3 to 2019:Q2, the majority of parameters in both models are estimated to be statistically significant. This includes most of the behavioral parameters to be found within the set of forecasting heuristics. A novelty of our analysis is that we are reliably able to identify the parameter for the intensity of choice in a macroeconomic model. Clearly statistically significant estimates between approximately 1.4 and 1.5 reveal a rather smooth and slow turnover from one forecasting heuristic to the other over time. The estimation results for the model under RE are also confirmed by applying the classical MLE. However, both models partially fail to capture the cross-correlation structure within the output-inflation nexus as the corresponding parameters turn out to be (close to) insignificant. This raises the issue of a potential model mis-specification under bounded and perfect rationality.

We conclude that this study considerably expands our understanding of heuristic switching models with heterogeneous agents used in macroeconomic research. Specifically, concerning the intensity of choice parameter, our empirical results are fruitful for the future parametrization of behavioral NKMs used for policy analysis. Future work in this area should focus on the estimation of more complex large-scale macroeconomic models.

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.
- Angeletos, G.-M. and C. Lian (2018). Forward guidance without common knowledge. *American Economic Review* 108(9), 2477–2512.
- Anufriev, M., T. Bao, and J. Tuinstra (2016). Microfoundations for switching behavior in heterogeneous agent models: An experiment. *Journal of Economic Behavior & Organization* 129, 74 – 99.
- Anufriev, M. and C. Hommes (2012). Evolutionary selection of individual expectations and aggregate outcomes in asset pricing experiments. *American Economic Journal: Microeconomics* 4(4), 35–64.
- Anufriev, M., C. Hommes, and T. Makarewicz (2018). Simple forecasting heuristics that make us smart: Evidence from different market experiments. *Journal of the European Economic Association* 17(5), 1538–1584.
- 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.
- Assenza, T., P. Heemeijer, C. Hommes, and D. Massaro (2013). Individual expectations and aggregate macro behavior. Tinbergen Institute Discussion Papers 13-016/II, Tinbergen Institute.
- Assenza, T., P. Heemeijer, C. Hommes, and D. Massaro (2019). Managing self-organization of expectations through monetary policy: A macro experiment. *Journal of Monetary Economics*.
- Barde, S. (2017). A practical, accurate, information criterion for Nth order Markov processes. *Computational Economics* 50(2), 281–324.
- Barde, S. (2020). Macroeconomic simulation comparison with a multivariate extension of the Markov information criterion. *Journal of Economic Dynamics and Control* 111, 103795.
- 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.
- Binder, M. and M. H. Pesaran (1999). *Multivariate Rational Expectations Models and Macroeconometric Modeling: A Review and Some New Results*, Chapter 3, pp. 111–155. John Wiley & Sons, Ltd.
- Boswijk, H. P., C. H. Hommes, and S. Manzan (2007). 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.
- Coibion, O. and Y. Gorodnichenko (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy* 120(1), 116–159.
- Coibion, O. and Y. Gorodnichenko (2015, August). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review* 105(8), 2644–78.
- 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.
- De Grauwe, P. and Y. Ji (2020). Structural reforms, animal spirits, and monetary policies. *European Economic Review* 124, 103395.
- Deák, S., P. Levine, J. Pearlman, and B. Yang (2017). Internal rationality, learning and imperfect information. Technical report, University of Surrey.
- Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007). On the fit of New Keynesian models. *Journal of Business & Economic Statistics* 25(2), 123–143.
- Dieci, R. and X.-Z. He (2018). Heterogeneous agent models in finance. In C. Hommes and B. LeBaron (Eds.), *Handbook of Computational Economics*, Volume 4 of *Handbook of Computational Economics*, Chapter 5, pp. 257–328. Elsevier.

- 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.
- Doucet, A., M. K. Pitt, G. Deligiannidis, and R. Kohn (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* 102(2), 295–313.
- Fagiolo, G., M. Guerini, F. Lamperti, A. Moneta, and A. Roventini (2019). *Validation of Agent-Based Models in Economics and Finance*, pp. 763–787. Springer International Publishing.
- Fernández-Villaverde, J. and P. Guerrón-Quintana (2020). Estimating DSGE models: Recent advances and future challenges. *NBER Working Paper* (w27715).
- Fernández-Villaverde, J. and J. F. Rubio-Ramírez (2007). Estimating macroeconomic models: A likelihood approach. *The Review of Economic Studies* 74(4), 1059–1087.
- Fernández-Villaverde, J., J. F. Rubio-Ramírez, and F. Schorfheide (2016). Solution and estimation methods for DSGE models. In *Handbook of Macroeconomics*, Volume 2, pp. 527–724. Elsevier.
- Franke, R. (2019, July). Bringing a baseline Harrod-Kaldor business cycle model to the data. Technical report, University of Kiel.
- 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.
- Franke, R. and F. Westerhoff (2017). Taking stock: A rigorous modelling of animal spirits in macroeconomics. *Journal of Economic Surveys* 31(5), 1152–1182.
- Galí, J. (2015). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications* (2nd ed.). Number 10495 in Economics Books. Princeton University Press.
- Gaunersdorfer, A. and C. Hommes (2007). *A Nonlinear Structural Model for Volatility Clustering*, pp. 265–288. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Giannoni, M., C. Patterson, and M. D. Negro (2015). The forward guidance puzzle. 2015 Meeting Papers 1529, Society for Economic Dynamics.
- Gigerenzer, G. and H. Brighton (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science* 1(1), 107–143.
- Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics* 8(S1), S85–S118.
- Goy, G., C. Hommes, and K. Mavromatis (2020). Forward guidance and the role of central bank credibility under heterogeneous beliefs. *Journal of Economic Behavior & Organization*.
- 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.
- Haldane, A. G. and V. Madouros (2012). The dog and the frisbee. Proceedings, Economic Policy Symposium, Jackson Hole, Federal Reserve Bank of Kansas City.
- Hamilton, J. D. (1994). State-space models. *Handbook of Econometrics* 4, 3039–3080.
- Hendry, D. and G. E. Mizon (2010). On the mathematical basis of inter-temporal optimization. Economics Series Working Papers 497, University of Oxford, Department of Economics.
- Herbst, E. and F. Schorfheide (2014). Sequential Monte Carlo sampling for DSGE models. *Journal of Applied Econometrics* 29(7), 1073–1098.
- Herbst, E. P. and F. Schorfheide (2015). *Bayesian estimation of DSGE models*. Princeton University Press.
- Hommes, C. (2013). *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems*. Cambridge University Press. Cambridge Books Online.
- Hommes, C., D. Massaro, and M. Weber (2019). Monetary policy under behavioral expectations: Theory and experiment. *European Economic Review* 118, 193–212.
- Hommes, C. H. (2006). 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*, Chapter 23, pp. 1109–1186. Elsevier.
- Hommes, C. H. (2021). Behavioral & experimental macroeconomics and policy analysis: A complex systems approach. *Journal of Economic Literature*, forthcoming.

- Jang, T.-S. (2012). *Structural estimation of the New-Keynesian model: A formal test of backward-and forward-looking behavior*. Emerald Group Publishing Limited.
- 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 (2019). Forecast heuristics, consumer expectations, and New-Keynesian macroeconomics: A horse race. *Journal of Economic Behavior & Organization*.
- Kleibergen, F. and S. Mavroeidis (2014). Identification issues in limited-information Bayesian analysis of structural macroeconomic models. *Journal of Applied Econometrics* 29(7), 1183–1209.
- Kristensen, D. (2009). 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. (2019). Simulated maximum likelihood estimation of agent-based models in economics and finance. In A. S. Chakrabarti, L. Pichl, and T. Kaizoji (Eds.), *Network Theory and Agent-Based Modeling in Economics and Finance*, pp. 203–226. Singapore: Springer Singapore.
- Kukacka, J. and J. Barunik (2017). Estimation of financial agent-based models with simulated maximum likelihood. *Journal of Economic Dynamics & Control* 85, 21–45.
- Kulish, M. and A. Pagan (2017). Estimation and solution of models with expectations and structural changes. *Journal of Applied Econometrics* 32(2), 255–274.
- Lamperti, F. (2018a, Apr). Empirical validation of simulated models through the GSL-div: An illustrative application. *Journal of Economic Interaction and Coordination* 13(1), 143–171.
- Lamperti, F. (2018b). 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.
- Lee, D. and K. Song (2015). Simulated maximum likelihood estimation for discrete choices using transformed simulated frequencies. *Journal of Econometrics* 187(1), 131–153.
- Lindé, J. (2005). Estimating New-Keynesian Phillips curves: A full information maximum likelihood approach. *Journal of Monetary Economics* 52(6), 1135–1149. The econometrics of the New Keynesian price equation.
- 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.
- Lux, T. and R. C. Zwinkels (2018). Empirical validation of agent-based models. In C. Hommes and B. LeBaron (Eds.), *Handbook of Computational Economics*, Volume 4, Chapter 8, pp. 437–488. Elsevier.
- 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.
- Simon, H. A. (1972). *The Sciences of the Artificial*. MIT Press.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica* 48(1), 1–48.
- Stock, J. H. and M. W. Watson (1999). Forecasting inflation. *Journal of Monetary Economics* 44(2), 293 – 335.
- Woodford, M. (2019). Monetary policy analysis when planning horizons are finite. *NBER Macroeconomics Annual* 33, 1–50.

Figures

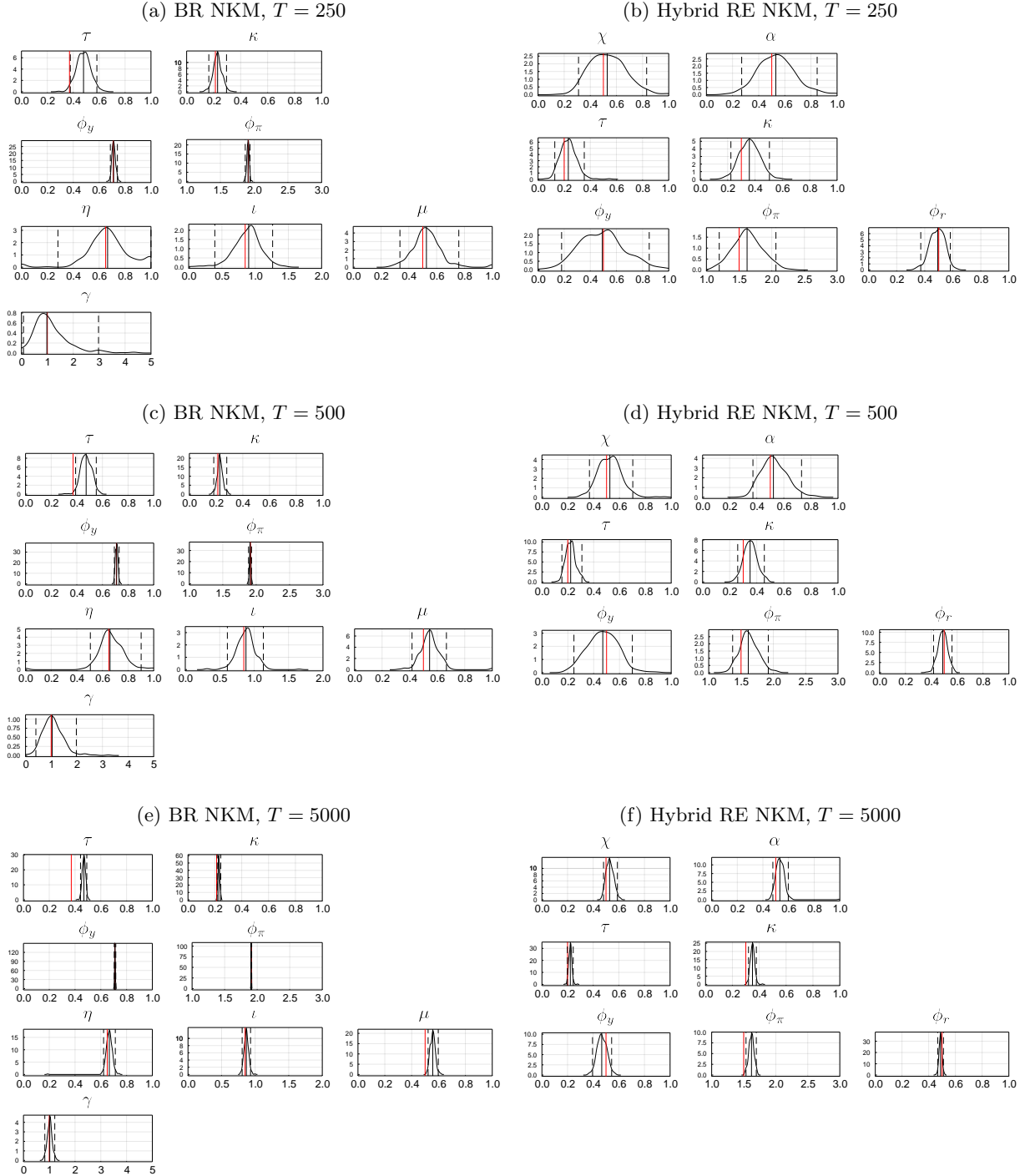


Figure 1: Densities of the pseudo-true parameter estimates. The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, and the dashed red vertical lines depict the 95% confidence intervals of the sample estimates. Based on 300 random runs, the parametrization follows [Table 1](#).

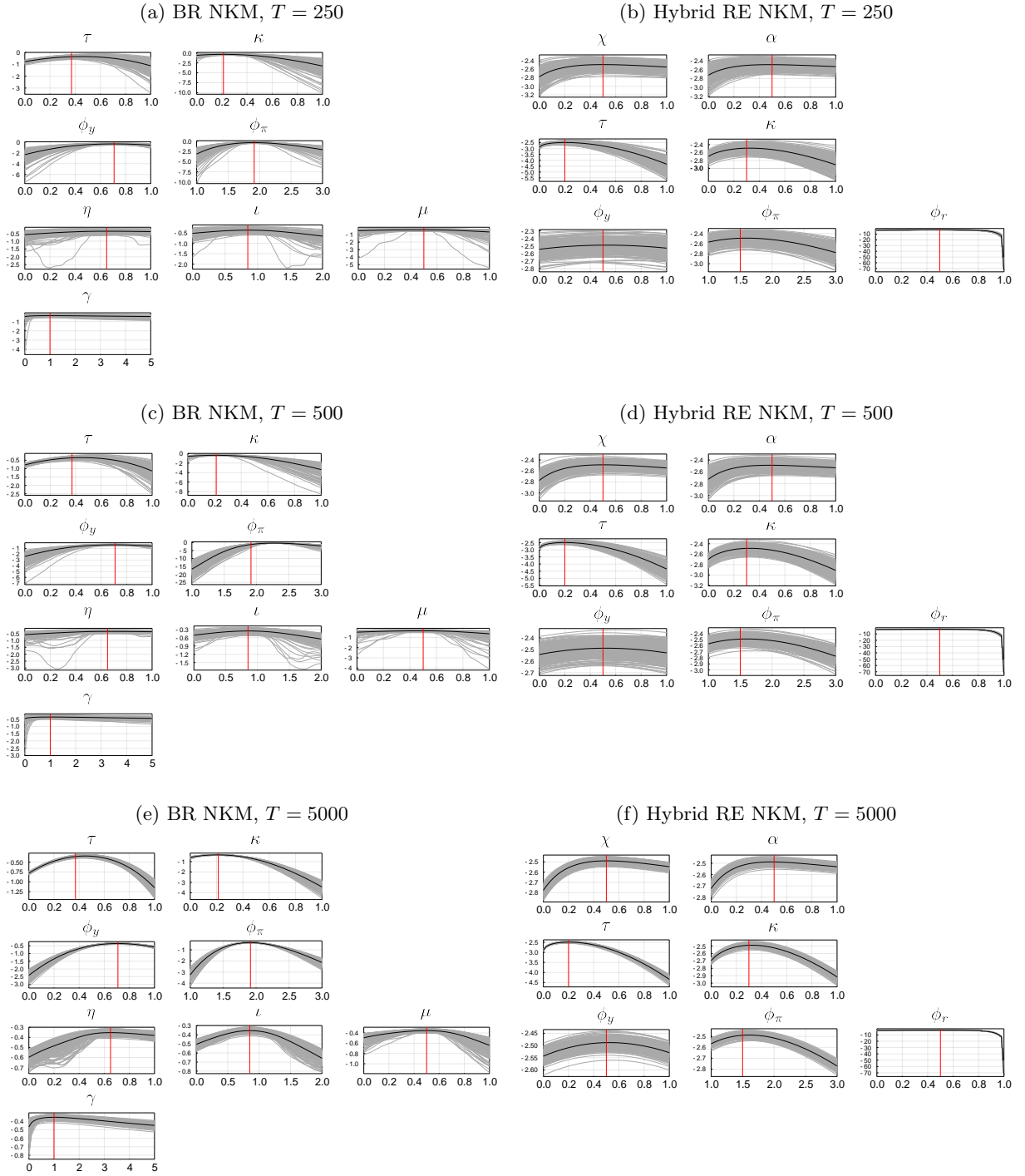


Figure 2: Profiles of the simulated log-likelihood function. The bold red vertical lines show the pseudo-true values, and the bold black curves depict the average log-likelihood. Based on 300 random runs, the parametrization follows [Table 1](#).

Appendix A: Robustness Analysis

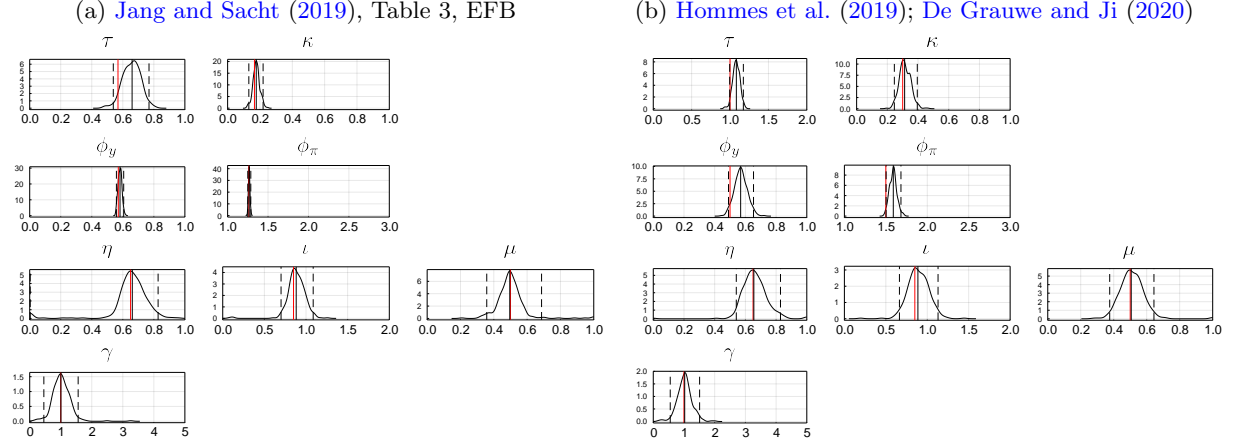


Figure 3: Densities of the pseudo-true parameter estimates for the BR NKM. The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, and the dashed red vertical lines depict the 95% confidence intervals of the sample estimates. Based on 300 random runs, $T = 500$, the parametrization of the structural parameters and the standard deviations of shocks are based on the cited papers, and the parametrization of the BR parameters follows Table 1.

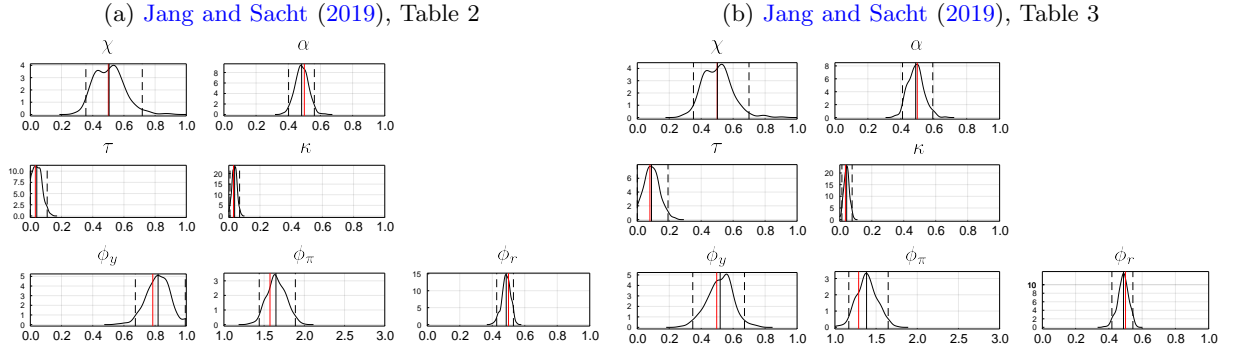
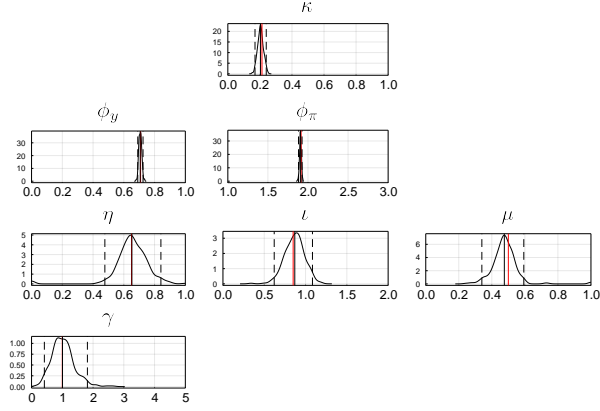
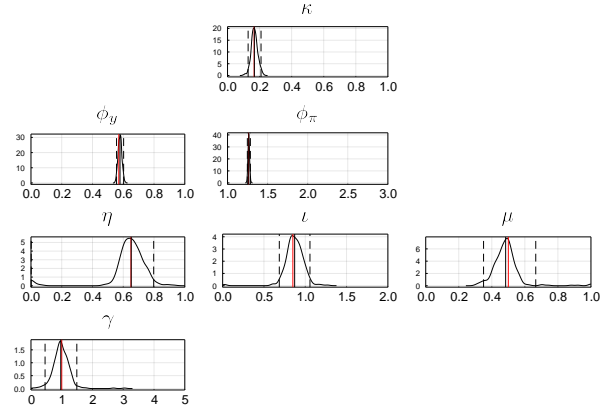


Figure 4: Densities of the pseudo-true parameter estimates for the hybrid RE NKM. The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, and the dashed red vertical lines depict the 95% confidence intervals of the sample estimates. Based on 300 random runs, $T = 500$, the parametrization of the intrinsic persistence parameters χ , α , ϕ_r and the standard deviations of shocks follow De Grauwe and Ji (2020), and the parametrization of the remaining structural parameters is based on Jang and Sacht (2019, Tables 2 and 3, REH scenarios).

(a) Parametrization follows [Table 1](#)



(b) [Jang and Sacht \(2019\)](#), Tab. 3, EFB



(c) [Hommes et al. \(2019\)](#); [De Grauwe and Ji \(2020\)](#)

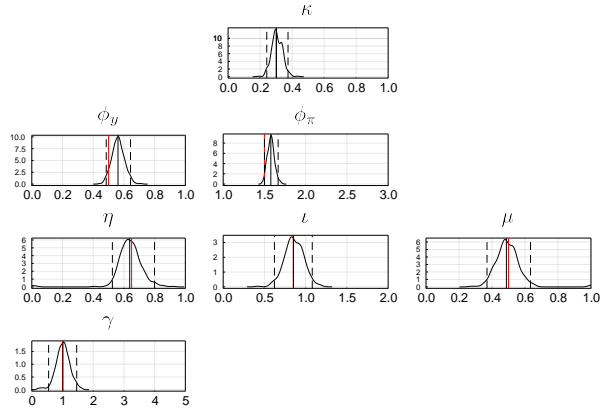


Figure 5: Densities of the pseudo-true parameter estimates for the BR NKM with fixed τ and alternative parametrizations. The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, and the dashed red vertical lines depict the 95% confidence intervals of the sample estimates. Based on 300 random runs, $T = 500$, the parametrization of the structural parameters and the standard deviations of shocks for panels (b) and (c) are based on the cited papers, and the parametrization of the BR parameters follows [Table 1](#).

Appendix B: Alternative Sets of Heuristics

In this section, we apply a robustness exercise with respect to alternative sets of forecasting heuristics. At its core, the model consists of the same structure as given by the equations (1)-(3) with $\chi = \alpha = \phi_r = 0$. The expressions for the market forecast (7) and the switching mechanism in general up to equation (9) remain unaltered but must be adjusted under the consideration of the heuristics displayed in the following Table 3.

Table 3: Alternative sets of forecasting heuristics

| Alternative set of heuristics | Types of agents | Paper |
|--|--|---|
| A1 $E^F y_{t+1} = \bar{y} + \psi_y(y_{t-1} - \bar{y})$ $E^C y_{t+1} = y_{t-1} + \xi_y(y_{t-1} - y_{t-2})$ | Fundamentalists Chartists | Gaunersdorfer and Hommes (2007) |
| A2 $E_{y_{t+1}}^O = (1/2)(\beta + \delta\lambda_{y,t})$ $E_{y_{t+1}}^P = -(1/2)(\beta + \delta\lambda_{y,t})$ | Optimists Pessimists | De Grauwe (2011) |
| A3 $E^f y_{t+1} = \bar{y}$ $E^N y_{t+1} = y_{t-1}$ | ‘Strict’ fundamentalists Naïve ones | De Grauwe and Ji (2020) |

The behavioral parameters ψ_y, ξ_y, β , and δ denote the speed of convergence, the degree of extrapolation, the subjective mean value, and the degree of divergence, respectively. The time-varying unconditional standard deviation in output is given by $\lambda_{y,t}$, where the steady state output gap is denoted by \bar{y} .

For an in-depth description of the alternative sets A1 and A2, we refer directly to [Jang and Sacht \(2019\)](#). With respect to set A3, it becomes apparent that $\psi_y = \xi_y = 0$ holds as we consider the heuristics included in set A1. Note also that in general, we assume zero values for the steady states in output and inflation, i.e., that $\bar{y} = \bar{\pi} = 0$ holds. Inflation expectations are formed consistently for A1-A3 based on two heuristics for the so-called extrapolators (E) and targeters (T):

$$E^E \pi_{t+1} = \pi_{t-1} \quad (19)$$

$$E^T \pi_{t+1} = \bar{\pi}. \quad (20)$$

Therefore, regardless of the alternative sets under consideration, inflation expectations are always formed based on the expressions (19) and (20) above. This is in line with the approaches discussed in [Jang and Sacht \(2019\)](#) and [De Grauwe and Ji \(2020\)](#). Figure 6 displays the outcome

of a Monte Carlo study for all alternative sets A1-A3 under a setup similar to [Section 5](#).

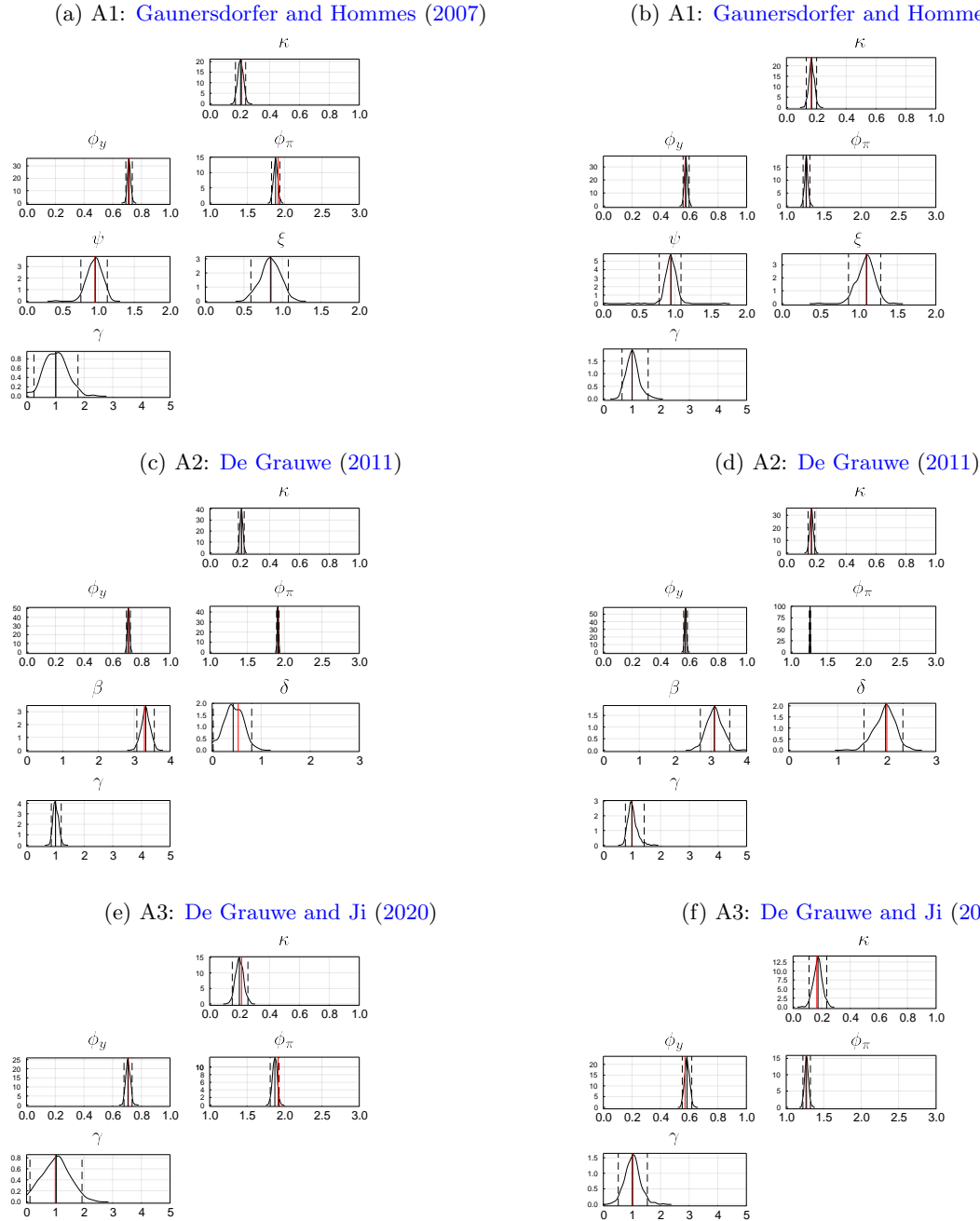


Figure 6: Densities of the pseudo-true parameter estimates for the BR NKM with alternative sets of heuristics. The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, and the dashed red vertical lines depict the 95% confidence intervals of the sample estimates. Based on 300 random runs, $T = 500$, the parametrization is based on [Jang and Sacht \(2019, Tables 2 \(left half\) and Table 3 \(right half\), EFB scenarios\)](#).

Appendix C: MLE for the Hybrid RE NKM

This section derives the classical MLE based on the true likelihood for the hybrid RE NKM. We take advantage of the linearity of the RE specification of the model and follow the MLE derivation in [Hamilton \(1994\)](#); [Lindé \(2005\)](#); [Jang \(2012\)](#); [Franke et al. \(2015\)](#).⁹ We also estimate the hybrid RE NKM via MLE as an additional empirical exercise, which is reported in the last column of [Table 2](#). We compare the results with those based on the SML estimator in [Subsection 6.3](#) and conclude that the SML approach approximates the MLE method very well.

Based on the state space notation in [sections 3.4](#) and [4](#), the law of motion follows:

$$\mathbf{X}_t = \mathbf{\Omega}\mathbf{X}_{t-1} + \mathbf{\Phi}\mathbf{\Gamma}_t, \quad \mathbf{\Gamma} \sim i.i.d. \mathcal{N}(0, \mathbf{\Sigma}_{\mathbf{\Gamma}}), \quad (21)$$

where $\mathbf{\Omega}$ and $\mathbf{\Phi}$ are the resulting RE-based projection matrices that can be obtained using a typical iterative method, and $\mathbf{\Sigma}_{\mathbf{\Gamma}} = \mathbf{I}(\sigma_y^2, \sigma_\pi^2, \sigma_r^2)'$, where \mathbf{I} is a 3×3 identity matrix. The conditional probability for the vector of the state variables \mathbf{X}_t is then

$$\mathbf{X}_t | \mathbf{X}_{t-1} \sim \mathcal{N}(\mathbf{\Omega}\mathbf{X}_{t-1}, \mathbf{\Phi}\mathbf{\Sigma}_{\mathbf{\Gamma}}\mathbf{\Phi}'). \quad (22)$$

The likelihood function looks like

$$L_T(\theta) = -\frac{3T}{2}\ln(2\pi) - \frac{T}{2}\ln|\mathbf{\Phi}\mathbf{\Sigma}_{\mathbf{\Gamma}}\mathbf{\Phi}'| - \frac{1}{2}\sum_{t=2}^T \mathbf{\Gamma}_t'(\mathbf{\Phi}\mathbf{\Sigma}_{\mathbf{\Gamma}}\mathbf{\Phi}')^{-1}\mathbf{\Gamma}_t \quad (23)$$

and the standard maximizer of the conditional log-likelihoods is already displayed in [equation \(15\)](#). Finally, the asymptotic properties of the MLE $\tilde{\theta}$ that allow for the estimation of the standard errors of the estimated coefficients lead to

$$\sqrt{T}(\tilde{\theta} - \theta) \sim \mathcal{N}\left(0, \left(\frac{\mathbf{H}}{T}\right)^{-1}\right), \quad (24)$$

where the Hessian information matrix is given by $\mathbf{H} = E\left[\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'}\right]$.

⁹We thank Tae-Seok Jang for sharing the Matlab code with us, which is needed to apply the MLE, and his related consultations on the MLE derivation. This code was written for use in [Franke et al. \(2015\)](#).

Online Appendices (not for publication)

Kukacka and Sacht; February 2021; Estimation of Heuristic Switching in Behavioral Macroeconomic Models

Online Appendix A: Important Properties of the SML Estimator

[Kristensen and Shin \(2012\)](#) argue that the main advantage of the SML 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 SML works even if the observables z_t are non-stationary. An important issue to consider is the potential curse of dimensionality with respect to the dimension of the vector of observables, as we smooth over only 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 SML for the three-equation NKM in [Section 5](#). Importantly, the SML does not suffer from the usual curse of dimensionality associated with kernel estimators, as substantially discussed in [Kristensen and Shin \(2012\)](#). The variance component of the resulting estimator does not need to be controlled by an unbearably large number of simulations, as the summation in equation (17) reveals an additional smoothing effect, and the additional variance of $\hat{L}_T(\theta)$ caused by simulations recovers the standard parametric rate of $1/N$. Therefore, the curse of dimensionality remains only of order $l \equiv \dim(z_t)$, and the SML will behave similarly to other estimation techniques including the MLE in this respect.

In contrast, given the kernel approximation method and its asymptotic properties, the simulated $\hat{L}_T(\theta)$ is generally a biased estimator of the actual $L_T(\theta)$ for a fixed approximation precision N and bandwidth $H > \mathbf{0}$. Only $N \rightarrow \infty$ and $H \rightarrow \mathbf{0}$ imply asymptotic consistency. Careful attention thus needs to be devoted to the selection of the bandwidth H with respect to the simulation size and a specific sample of data. Fortunately, in a simulation study, [Kristensen and Shin \(2012\)](#) demonstrate that the SML performs well using a broad range of bandwidths. A standard identification assumption for the stationary case requires $\mathbb{E}[\log p(z_t|x_t, \theta)] < \mathbb{E}[\log p(z_t|x_t, \theta_0)]$, $\forall \theta \neq \theta_0$. [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} as $N \rightarrow \infty$. This suggests that the proposed methodology is robust to the choice of H from a theoretical perspective because one can assuredly well identify the model parameters in large finite samples after \bar{H} is set. However, in a practical application, one still needs to know the threshold level of \bar{H} that can be examined through simulations. In addition to a proper selection of N and H , the kernel K itself is assumed to satisfy conditions K.1–2 specified in [Kristensen and Shin \(2012, p. 81\)](#), i.e., to be continuously differentiable, allow for unbounded support, and belong to so-called higher-(than second)-order or bias-reducing kernels. For instance, from the most commonly used kernels, the Gaussian kernel naturally satisfies all given assumptions. A higher number of derivatives of p 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 SML $\hat{\theta}$ is first-order asymptotically equivalent to the infeasible MLE $\tilde{\theta}$ under a set of general conditions satisfied by most models and ensures that $\hat{p} \rightarrow p$ sufficiently fast, which even allows for mixed discrete and continuous distributions and the non-stationarity of the dependent variables. A set of regularity conditions (A.1–4, K.1–2, p. 80–81) on the model and its associated conditional density is defined to satisfy these general conditions for the uniform convergence rates of the kernel estimators stated in [Kristensen \(2009\)](#). Moreover, under additional assumptions including, e.g., stationarity, the results regarding the higher-order asymptotic properties together with the expressions for the bias and variance components of the SML estimator due to kernel estimation and numerical simulations are derived.

Online Appendix B: Memory Parameter

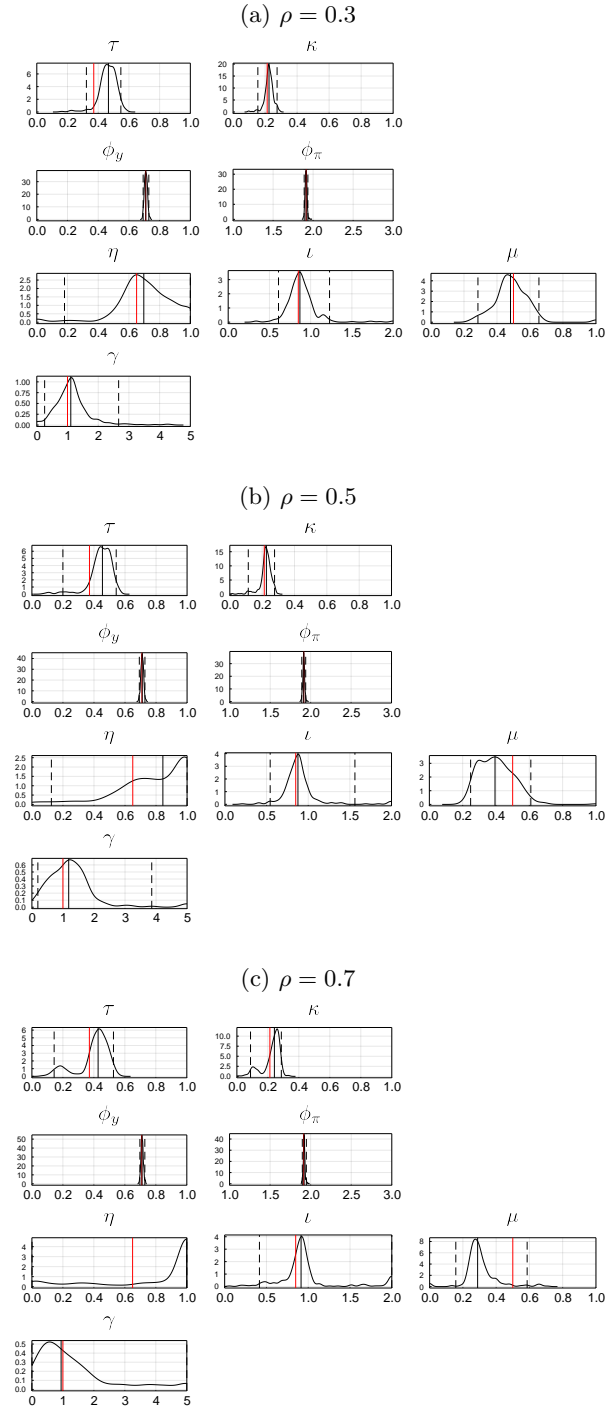


Figure 7: Densities of the pseudo-true parameter estimates for an alternative BR NKM with a non-zero memory parameter ρ . The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, and the dashed red vertical lines depict the 95% confidence intervals of the sample estimates. Based on 300 random runs, $T = 500$, the parametrization follows [Table 1](#).

Online Appendix C: Comparison to the Simulated Method of Moments

Table 4: Estimates for the BR NKM via the SMM

| | Par. | BR forward-looking NKM | | |
|------------------------|--------------|------------------------|-------------|-------------|
| | | $T=250$ | 500 | 5000 |
| τ | <i>.371</i> | .38 | .37 | .38 |
| $\langle 0, 1 \rangle$ | | (.26-.50) | (.29-.47) | (.32-.44) |
| κ | <i>.213</i> | .21 | .22 | .21 |
| $\langle 0, 1 \rangle$ | | (.14-.32) | (.16-.30) | (.17-.29) |
| ϕ_y | <i>.709</i> | .70 | .71 | .71 |
| $\langle 0, 1 \rangle$ | | (.52-.94) | (.56-.90) | (.55-.87) |
| ϕ_π | <i>1.914</i> | 1.94 | 1.95 | 1.97 |
| $\langle 1, 3 \rangle$ | | (1.65-2.25) | (1.73-2.17) | (1.77-2.20) |
| η | <i>.65</i> | .72 | .77 | .62 |
| $\langle 0, 1 \rangle$ | | (.02-.99) | (.01-.99) | (.07-.98) |
| ι | <i>.85</i> | .69 | .64 | .70 |
| $\langle 0, 2 \rangle$ | | (.13-1.23) | (.16-1.08) | (.28-1.08) |
| μ | <i>.50</i> | .54 | .51 | .54 |
| $\langle 0, 1 \rangle$ | | (.23-.79) | (.27-.74) | (.40-.68) |
| γ | <i>1.00</i> | 1.96 | 2.05 | 2.13 |
| $\langle 0, 5 \rangle$ | | (.04-4.72) | (.05-4.70) | (.08-4.64) |

The constraints for optimization and its starting point are given in $\langle \rangle$ brackets. T denotes the length of the executed time series. The sample medians based on 300 random runs are reported, while the 95% confidence intervals of the sample estimates are reported in $()$ parentheses. The parametrization follows [Table 1](#). The figures are rounded to 2 or 3 decimal places. The SMM setup follows [Jang and Sacht \(2016, 2019\)](#); [Franke \(2019\)](#). The reported results are directly comparable to the left half of [Table 1](#).

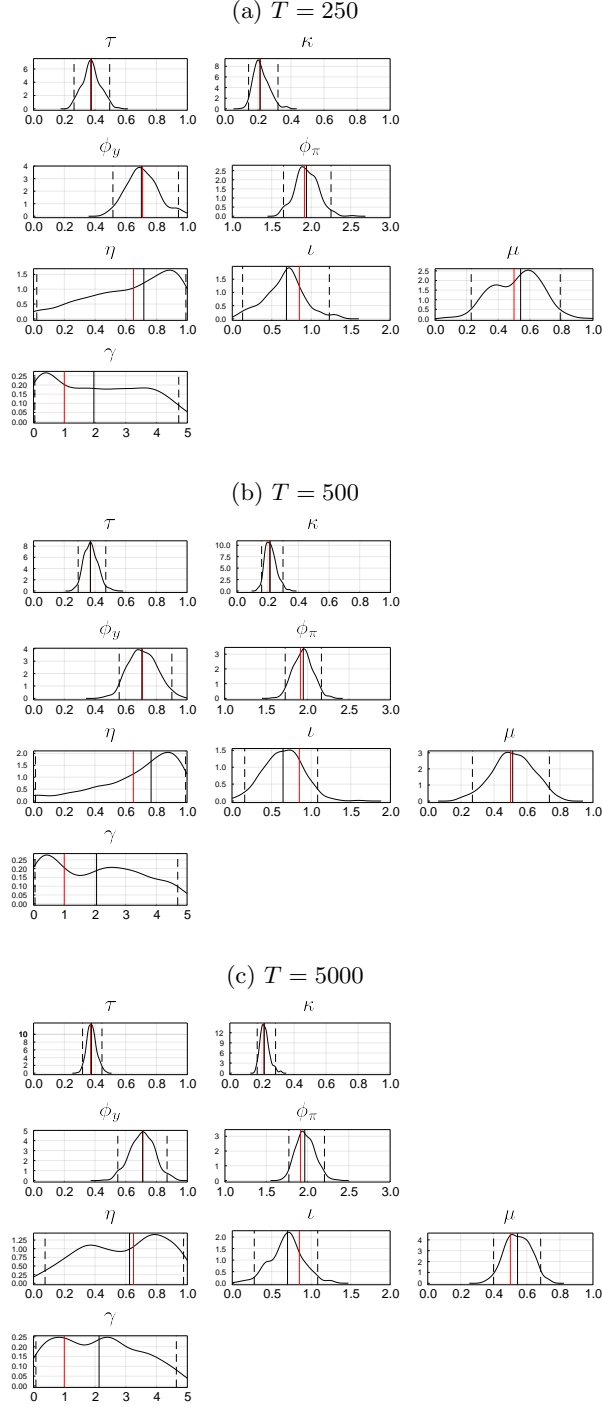


Figure 8: Densities of the pseudo-true parameter estimates for the BR NKM via the SMM following the setup by [Jang and Sacht \(2016, 2019\)](#); [Franke \(2019\)](#). The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, and the dashed red vertical lines depict the 95% confidence intervals of the sample estimates. Based on 300 random runs, the parametrization follows [Table 1](#). The reported results are directly comparable to the left half of [Figure 1](#).