

Lux, Thomas

Working Paper

Lack of identification of parameters in a simple behavioral macroeconomic model

Economics Working Paper, No. 2024-02

Provided in Cooperation with:

Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Lux, Thomas (2024) : Lack of identification of parameters in a simple behavioral macroeconomic model, Economics Working Paper, No. 2024-02, Kiel University, Department of Economics, Kiel

This Version is available at:

<https://hdl.handle.net/10419/300523>

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.

Lack of Identification of Parameters in a Simple Behavioral Macroeconomic Model

by Thomas Lux



Lack of Identification of Parameters in a Simple Behavioral Macroeconomic Model *

Thomas Lux [†]

Department of Economics, University of Kiel, Germany
and

Department of Economics, University Jaume I Castellón, Spain

January 8, 2024

Abstract

Identifiability of the parameters is an important precondition for consistent estimation of models designed to describe empirical phenomena. Nevertheless, many estimation exercises proceed without a preliminary investigation into the identifiability of its models. As a consequence, the estimates could be essentially meaningless if convergence to the ‘true’ parameters is not guaranteed in the pertinent problem. We provide some evidence here that such a lack of identification is responsible for the inconclusive results reported in recent literature on parameter estimates for a certain class of nonlinear behavioral New Keynesian models. We also show that identifiability depends on the subtle details of the model structure. Hence, a careful investigation of identifiability should precede any attempt at estimation of such models.

JEL Classification: C53, E12, E32

Keywords: Behavioral macro, identification, forecast heuristics

*Helpful comments by Daniel Fehrle, Stephen Sacht, Leonardo Bargigli, Giorgio Ricchiuti and participants of the 2nd DISEI workshop on Heterogeneity, Evolution and Networks in Economics are gratefully acknowledged. Financial support by the Generalitat Valenciana under the project AICO/2021/005 is gratefully acknowledged.

[†]Author’s Address: Department of Economics, University of Kiel, Olshausenstr. 40, 24118 Kiel, Germany, Email: lux@economics.uni-kiel.de.

1 Introduction

Over the last two decades, models with certain deviations from completely rational behavior have more and more become accepted in the economics profession. Experimental economics has unearthed a plethora of behavioral ‘anomalies’ that are not easily reconcilable with perfectly rational optimization and a large spectrum of ‘behavioral’ theories has been put forward to explain various phenomena and stylized facts (e.g., Thaler, 2015; Akerlof and Shiller, 2010, to only quote Nobel laureates from the behavioral community).

In recent times, behavioral components have also been incorporated into otherwise relatively traditional macroeconomic models of the New Keynesian tradition, and new models have been proposed that attempt to reconstruct macroeconomics from a bottom-up approach based on the economic interactions of a large pool of agents. Prominent examples of the first line of approach are the contributions by Anufriev et al. (2013), Branch and McGough (2009, 2010), de Grauwe and co-authors (e.g., de Grauwe, 2011, 2012; de Grauwe and Ji, 2020) and Lines and Westerhoff (2010, 2012), among others. Leading examples of the second school are delli Gatti et al. (2011) or Poledna et al. (2023).

The development of this new class of models has also brought the task of their empirical validation to the fore. In order to compete on the same footing with tradi-

tional Dynamic Stochastic General Equilibrium (DSGE) models, rigorous estimation methods are required as well as appropriate tools for model selection in the presence of a variety of different specifications of both traditional and behavioral macro models. Some research in this vein is already available with a bunch of papers exploring various estimation methods for different versions of the de Grauwe (2012) version of the behavioral New Keynesian (NKM) model (Grazzini et al., 2017; Jang and Sacht, 2016, 2021; Kukačka and Sacht, 2023), while fewer attempts at empirical validation of more complex ones exist (e.g., delli Gatti and Grazzini, 2020).

Proper estimation requires certain conditions to be met by the model under scrutiny, often called ‘regularity conditions’ in statistics and econometrics. Some of these are easy to verify and hold for large classes of models. For example, most behavioral agent-based models are Markovian despite their often involved interactions between agents, a feature that goes a long way already towards satisfaction of certain sets of regularity conditions. A necessary condition that is needed in virtually every validation exercise is identifiability of the parameters. Identification issues have received some attention in the empirical DSGE literature. Canova and Sala (2009) were the first to address problems of identification of certain parameters of DSGE models, due to observational equivalence in the mapping of the structural parameters to the coefficients of linear DSGE models. Conditions for identification of the parameters of log-linearized DSGE models have been established later on by Iskrev (2010) and

applied in a computational manner (checking satisfaction of the conditions for a large set of parameter values) to the Smets and Wouters (2007) model.

One might hope that nonlinearity could resolve certain identification problems as strict collinearity¹ of two or more parameters would at most hold locally in a nonlinear model when the two parameters enter in different ways in its nonlinear functional components. From this example, one might hope that identification is to a certain extent generic in models of the highly nonlinear format of most behavioral or agent-based macro models.

There are, however, hardly any general results on conditions for identification of parameters in nonlinear models. McManus (1992) has demonstrated that global identification is generic in a certain class of structural models with exogenous variables, but otherwise, rigorous knowledge on identification of other model classes is virtually non-existent. While the nonlinearity of behavioral models might make identification of parameters a plausible property, it also, unfortunately, impedes a straight forward formal analysis of the identifiability issue. As a consequence, practically all the available empirical work on validation of such models simply assumes identifiability or does not even mention this issue. As it turns out, however, identifiability of its parameters might indeed constitute a serious problem for certain behavioral macro models. Experimenting with different versions of the de Grauwe (2012) framework and different estimation approaches, we virtually always encountered problems of non-convergence

¹and, therefore, non-identifiability.

to the ‘true’ values when conducting Monte Carlo simulations to assess the accuracy of the estimates. Using a simple version of this model that had been estimated before by Grazzini et al. (2017) and Zhang et al. (2023) we provide a detailed analysis of this problem and try to shed light on the structural origin of the lack of identification in this highly nonlinear set-up.

To explore the issue of identifiability of parameters we use a Markov Chain Monte Carlo (MCMC) approach in this paper applying the adaptive adjustment of the proposals that had already been adopted for financial agent-based models in Lux (2022). As pointed out by Siekmann et al. (2012), MCMC can shed light on whether certain parameters of a model are identified or not, and what particular obstacles exist to their identification.

As a Bayesian approach, MCMC does not provide a point estimate of a parameter, but an approximation of the parameters’ posterior given certain priors. Various deficiencies could emerge in the attempt to implement this approach: The posterior could settle down with its support being concentrated in the regions far off the ‘true’ parameter values, or the posterior could show little difference from the priors which would indicate a lack of informativeness of the data with respect to the parameters. Essentially, we find that many of the behavioral parameters of the New Keynesian model of de Grauwe (2012) would require unrealistically large sample sizes to be identifiable with any satisfactory level of precision. For sample sizes typically available in

macroeconomics, we find that the posteriors do not show much difference from the priors, and for certain parameters, this even remains so for extremely large, unrealistic sample sizes (e.g., 20,000 observations). In single MCMC runs, the posterior would often appear to be concentrated away from the values that have been used in the simulation, while the degree of the mixing of the Markov chain (i.e., the number of accepted draws) is close to its theoretically optimal level. However, running multiple chains with different initial conditions shows that such seemingly good performance might be illusory reflecting the hovering of the chain within some limited neighbourhood of a very flat likelihood function.

Somewhat surprisingly, the severity of these problems might differ a lot between versions of the behavioral NKM that only differ slightly in their assumptions. In sec. 4, a small variation of the model of de Grauwe (2011, 2012) and de Grauwe and Ji (2020) is presented for which all parameters seem to be identifiable with sufficiently large samples although in this case as well, typical macroeconomic sample sizes do not lead to major changes of the posterior compared to the prior. The difference between both version boils down essentially to a somewhat less complex interaction of the parameters of the reduced form of the model in terms of the lagged variables and the propagation of shocks.

The plan of the remainder of this paper is as follows: Sec. 2 presents the behavioral New Keynesian model proposed by de Grauwe (2012) for which meanwhile a certain

body of literature exists that attempts its empirical validation. Sec. 3 shows the results of Monte Carlo simulations using a Markov Chain Monte Carlo estimation that provides evidence for a lack of identifiability of a model specification that has been used by a number of other authors before. As it turns out, identification of the behavioral parameters is at least very weak for reasonable sample sizes. While there is some improvement for certain parameters with unrealistically large samples, other parameters show no sign of convergence even for samples that exceed empirical records by two orders of magnitude. Sec. 3 also discusses the possible origin of such mis-specification. Sec. 4 shows that identification hinges on the precise details of the model structure: A very closely related specification of the behavioral NKM with only seemingly minor changes in the specification of the Taylor rule apparently does suffer to a lower degree from the issues highlighted for the original specification. Sec. 5 estimates this version of the model for a long series of quarterly U.S. macro data. As it turns out, this exercise does not provide much evidence for the relevance of the specific behavioral elements that are characteristic of the present variant of behavioral macro models. Sec. 5 concludes.

2 The Behavioral NKM model

We adopt the version of de Grauwe (2012) of a model of heuristic expectation formation within an otherwise very conventional New Keynesian setting. The New

Keynesian core of the model consists of its equations for the evolution of the output growth rate (g), the inflation rate (π) and the interest rate (r):

$$\begin{aligned} g_t &= a_1 E g_{t+1} + (1 - a_1) g_{t-1} + a_2 (r_t - E \pi_{t+1}) + \epsilon_{1,t} \\ \pi_t &= b_1 E \pi_{t+1} + (1 - b_1) \pi_{t-1} + b_2 g_t + \epsilon_{2,t} \\ r_t &= c_1 (\pi_t - \pi^*) + c_2 g_t + c_3 r_{t-1} + \epsilon_{3,t} \end{aligned} \tag{1}$$

In system (1), the first equation is the IS relation with partially forward looking behavior (E denoting expectations which are not necessarily ‘rational’ ones) and partially backward looking dependency (habit persistence being one possible explanation of such a structure). The second equation is the expectation adjusted Phillips curve which (sluggish price adjustment serving as a potential explanation here) also comes with a forward looking and a backward looking term. The third equation, finally, is the Taylor rule. Parameters $a_1, a_2, b_1, b_2, c_1, c_2, c_3$ are all expected to be positive except for a_2 which gives the usually negative effect of higher real interest rates on consumption and investment. $\epsilon_{1,t}$, $\epsilon_{2,t}$ and $\epsilon_{3,t}$ are all random variables which in empirical applications are assumed to be drawn from independent Normal distributions with means all equal to zero and variances σ_1^2 , σ_2^2 , σ_3^2 . When assuming rational expectations, the forward looking terms in eqs. (1) would be fully endogenous, and this relatively simple model would already leave us with 10 parameters to be estimated.

In de Grauwe (2012), the expectations are formed according to a popular discrete-

choice formalization allowing the agents to select from two simple alternatives: They either expect the fundamentals to govern the behavior of the macro variables and predict their equilibrium values for next period, or use the last observation available, i.e., g_{t-1} or π_{t-1} as forecasts for time $t + 1$. It is easy to see that when setting (normalizing) the target value for the inflation rate as $\pi^* = 0$, the only steady state of the system is $(g^*, \pi^*, r^*) = (0, 0, 0)$. Hence, the predictions are either $E^f g_{t+1} = 0$ or $E^e g_{t+1} = g_{t-1}$ as well as $E^f \pi_{t+1} = 0$ or $E^e \pi_{t+1} = \pi_{t-1}$. What fraction of agents chooses which forecast is determined by the fitness of the predictions f (for fundamentals) and e (for extrapolation):

$$\begin{aligned} P_{f,t}^g &= \frac{\exp(10\gamma U_{f,t})}{\exp(10\gamma U_{f,t}) + \exp(10\gamma U_{e,t})}, P_{e,t}^g = 1 - P_{f,t}^g \\ P_{f,t}^\pi &= \frac{\exp(10\gamma W_{f,t})}{\exp(10\gamma W_{f,t}) + \exp(10\gamma W_{e,t})}, P_{e,t}^\pi = 1 - P_{f,t}^\pi \end{aligned} \quad (2)$$

with the superscript indicating the two variables growth rate g and inflation π for which predictions are formulated, and the subscript denoting the fundamentalist and extrapolative predictions at time t^2 .

The four fitness functions are determined as loss functions.

$$\begin{aligned} U_{j,t} &= \rho U_{j,t-1} - (E^j g_{t-1} - g_{t-1})^2 \\ W_{j,t} &= \rho W_{j,t-1} - (E^j \pi_{t-1} - \pi_{t-1})^2 \end{aligned} \quad (3)$$

²The factor 10 in the exponential function follows previous authors and allows us to restrict the prior for the parameter γ to the interval $[0, 1]$.

with $U_{j,t}$, $W_{j,t}$ the fitness functions of prediction j ($j = f, e$) for the growth rate and the inflation rate, respectively. Note that for a plausible sequence of predictions, the functions $P_{j,t}^g$ and $P_{j,t}^\pi$ could last have been updated after the time $t-1$ realizations of g and π have been observed. The predictions Eg_{t-1} and $E\pi_{t-1}$ would then have been formulated in period $t-2$ and would have used the most recent observations then, i.e. g_{t-3} and π_{t-3} .

The system of equations (1) to (3) is obviously a highly non-linear dynamic system with stochastic components. It seems interesting to clarify more precisely the structural form of this system as to be able to refer to appropriate mathematical results on the behaviour of estimates of its parameters. Note that while eqs. (1) have only lags of order one, eqs. (3) introduce lags of order 3 in growth and inflation rates in the extrapolative heuristic. Higher order lags can be transformed into first-order lags by introducing auxiliary variables such as: $\hat{g}_t = g_{t-1}$, $\tilde{g}_t = \hat{g}_{t-1}$, $\hat{\pi}_t = \pi_{t-1}$, $\tilde{\pi}_t = \hat{\pi}_{t-1}$. Hence, we can write the system as a first-order nonlinear system with eleven dynamic variables: $g_t, \hat{g}_t, \tilde{g}_t, \pi_t, \hat{\pi}_t, \tilde{\pi}_t, r_t, U_{e,t}, U_{f,t}, W_{e,t}, W_{f,t}$. We can further decompose the variables into the observed and the latent ones writing the system's law of motion as

$$\begin{aligned} Y_{t+1} &= f(Y_t, X_t) + N(0, \Sigma), \\ X_{t+1} &= g(X_t, Y_t) \end{aligned} \tag{4}$$

with $Y_t = (g_t, \hat{g}_t, \tilde{g}_t, \pi_t, \hat{\pi}_t, \tilde{\pi}, r_t)$ collecting the observable variables and $X_t = (U_{f,t}, U_{e,t}, W_{f,t}, W_{e,t})$ the latent ones. The stochastic term, the multivariate Gaussian $N(0, \Sigma)$ in the first equation has non-zero variances only for g_t, π_t and r_t , but the covariance matrix has entries that also depend on the parameters of the contemporaneous effects in eqs. (1). A system of this type is denoted an *observation-driven system*. A typical example in the econometrics literature of a model with this structure is the seminal GARCH model for the time evolution of the second moment of financial assets. Note that the characteristic features of such a system are (i) stochasticity in the observed variable, but not in the latent ones, (ii) feedback effects from the observed to the latent variables.

With stochastic elements present in the latent part as well, one would rather speak of a general state-space model. Without feedback from the observed to the latent variables, the noise could be interpreted as measurement noise since it would not affect the hidden process X_t . Inspection of many behavioral or agent-based models shows that they often share the particular structure of eqs. (4) (cf. Lux, 2021). A series of recent papers has investigated the statistical properties of maximum likelihood estimation for observation-driven dynamic systems: Douc et al. (2013, 2015) demonstrates consistency of maximum likelihood for models with first-order lag structures under certain regularity conditions, while Sim et al. (2021) generalize these results to higher-order lag structures. However, the regularity conditions of

these papers imply identifiability of the parameters, which might not be guaranteed in practice.

In light of such promising theoretical behavior of general observation-driven models, recent Monte Carlo results reported in Zhang et al. (2023) on the above model appear quite sobering. The authors report on two Monte Carlo studies for the estimation of a 9-parameter version of the model (with c_1, c_2 and c_3 fixed) using both a sequential Monte Carlo (SMC) and a Markov-Chain Monte Carlo (MCMC) approach with a kernel-density approximation of the likelihood. They simulate the de Grauwe (2012) model with $T = 50, T = 300$ and $T = 400$ observations (their Tables 3 and 4). While the authors interpret their results as evidence for the potential of these approaches to recover the parameters of the model, a closer inspection of the reported results shows that for most parameters, the estimation accuracy *deteriorates* with sample size (for SMC) or the accuracy appears virtually unaffected by sample size (for MCMC). Both observations speak against consistent estimation of the parameters of the model. I will show in the following that the likely reason is a lack of identification of the parameters which indeed might unfortunately be more widespread for behavioural models of a similar format than previously thought.

3 Monte Carlo Simulations

We follow Grazzini et al. (2017) and Zhang et al. (2023) by adopting an MCMC approach for the estimation of the parameters of the de Grauwe (2012) model. In contrast to the previous applications, we do not, however, use any approximation to the likelihood, but the conditional likelihood itself, i.e.,

$$L_{\theta,T} = \prod_{t=2}^T P_{\theta}(Y_t|Y_{t-1}) \quad (5)$$

with T : the length of the pseudo-empirical time series whose parameters we wish to estimate, $Y_t = (g_t, \pi_t, r_t)$ the observable parameters (including, of course, their lags), $P_{\theta}(Y_t|Y_{t-1})$ the conditional density of the observations at time t under the model with parameter set θ , and the parameter set being $\theta = (a_1, a_2, b_1, b_2, \sigma_1, \sigma_2, \sigma_3, \rho, \gamma)$ while the parameters of the Taylor rule (c_1, c_2, c_3) have been assumed to be known.

According to the quoted results of Douc et al. (2013, 2015) we have some reason to assume that maximum likelihood is consistent for this model, provided the parameters are uniquely identified. We could, then, also be confident that using the likelihood in an MCMC setting, would lead to convergence to the posterior distribution. Denoting the prior distribution of the parameters by $p(\theta)$, and the proposal distribution by $g(\theta_{\xi}|\theta_{\xi-1})$ with ξ the sequential order of the generated Markov chain, we accept new

proposals θ^* with the probability

$$\alpha(\theta^*, \theta_{\xi-1}) = \min\left\{\frac{L_{\theta^*, TP}(\theta^*)g(\theta_{\xi-1}|\theta^*)}{L_{\theta_{\xi-1}, TP}(\theta_{\xi-1})g(\theta^*|\theta_{\xi-1})}, 1\right\} \quad (6)$$

For the prior distribution, we have used either the distributions for the different parameters used by Zhang et al. (2023) before (mostly adopted from the DSGE literature) or we have used mostly uniform priors over the same support as the more traditional ones. Which set we used did not make much of a difference, at least for the experiments with longer pseudo-empirical time series (as it should be as then the added information from the time series should dominate the influence of the prior). In our presentation of the Monte Carlo simulation runs below, we concentrate on results based on uniform priors for all parameters as these nicely allow to infer how much the posterior moves away from the prior, i.e., how informative the data are for the parameters. The more traditional (strong) priors used in Grazzini et al. (2017) and Zhang et al. (2023) come with certain restrictions of the parameters, $0 \leq a_1 \leq 3, a_2 \leq 0, 0 \leq b_1, b_2 \leq 1, 0 \leq \gamma, \rho \leq 1$, which we also mostly keep when using the uniform priors for these variables. For the distribution of the proposals, we used a fat-tailed multivariate Student t distribution with mean equal to the last accepted set of parameters, degrees of freedom equal to 3 and the counterpart of the variance-covariance matrix being adaptively adjusted on the base of the accepted previous draws of the chain.

The degree of adaptation is assumed to decrease with the length of the chain which guarantees that adaptation does not get in the way of convergence to the posterior (cf. Andrieu and Thomas, 2008). Adaptation of the proposal distribution in the MCMC algorithm serves to guarantee a sufficient degree of mixing of the chain without any necessary adjustment by hand, or trial and error. Results reported below demonstrate that this algorithm indeed meets its target acceptance rate very precisely. The details of the algorithm are explained in Appendix A.

This MCMC experiment is conducted for pseudo-empirical time series with length $T_1 = 200, T_2 = 2,000$ and $T_3 = 20,000$, the latter obviously going far beyond the size of any real-life macroeconomic sample. As in Zhang et al. (2023), we fix all parameters at 0.5 except for $a_2 = -0.5$ and $c_1 = 1.5$. For the dynamic coefficients of the three NKM equations of eqs. (1) this indeed amounts to values that are roughly representative of estimates for models with rational expectations. The standard deviations of the random terms in eqs. 1 are all set equal to 0.1. The three parameters of the Taylor rule are assumed to be known so that nine parameters remain to be estimated, i.e. the set $\theta = \{a_1, a_2, b_1, b_2, \rho, \gamma, \sigma_1, \sigma_2, \sigma_3\}$. In what we document below we have assumed uniform priors with support over $[0, 1]$ for all behavioral parameters except a_2 for which a uniform distribution over $[-1, 0]$ has been used. For the standard deviations $(\sigma_1, \sigma_2, \sigma_3)$ uniform priors over $[0, 2]$ have been used. Results did, however, not change significantly under more informative priors.

As far as we can see, it has never been explored whether models like the one studied here would show nice convergence properties if only enough data were available. Numerical explorations of the model of de Grauwe (2012) indicate that it exhibits completely ‘harmless’ behavior in its ‘deterministic skeleton’ (i.e. with stochastic elements switched off). For the parameter set used here as well as for all others we tried (even extreme ones), there was stable convergence to the unique equilibrium $g^* = \pi^* = r^* = 0$, even with very fast convergence over very few periods after shocks and irrespective of whether the Taylor principle ($c_1 > 1$) of rational expectations models was satisfied. No limit cycles, chaotic behavior or other ‘interesting’ dynamics were ever observed, which have often been found to be characteristic features of similar models with behavioral dynamics. Given this convenient behavior, we might have even more reason to expect the model to behave well under attempts of its empirical validation.

Table 1 and Figs. 1 through 3 show statistics for the posterior and graphical illustrations of the Markov chains for the three different sample sizes. In each case, we run the chain 10 times using different initial conditions over 2,001,500 steps in each case with the first 1,500 realizations obtained for warm up with a given variance-covariance matrix of proposals θ^* composed of all diagonal elements equal to 0.1². Afterwards, adaptation has been initiated for the variance-covariance matrix of the chain as explained in Appendix A. We can see from Table 1 that the acceptance rates

are close to what is seen as optimal in MCMC (the adaptation targets an acceptance rate of 0.234 as recommended in the literature). For the posteriors, the mean over the set union of the last 1,000,000 data points of all ten sequences is given together with the 95 percent credible interval, i.e. the 2.5 and 97.5 quantiles of the realizations of the chain. Also shown is a popular diagnostic statistic for the convergence of the Markov chain, the so called *potential scale reduction factor* \hat{R} proposed by Gelman and Rubin (1992). This measure is based on a comparison of the variance of the posteriors within samples and between samples, and is constructed in a way to be limited from below by unity. A large scale reduction factor indicates insufficient convergence signaling that longer simulations would be needed for convergence to the target distribution, or that single runs have not sufficiently explored the target distribution, and, therefore, their variances differ substantially from each other. Since \hat{R} is a stochastic quantity, the table also shows its upper 95 percent bound (denoted R_{95}) computed on the base of the approximate distribution of \hat{R} derived by Flegal et al. (2008). A typical recommendation is that \hat{R} should be below 1.1 for a Markov chain to exhibit satisfactory convergence. In Figs. 1 to 3, we provide illustrations of the development of the Markov chains of the behavioral parameters.

The results are dramatically poor: For the smallest sample size, $T_1 = 200$, the confidence diagnostic is often astronomically high for all behavioral parameters, indicating that Markov chains of much longer sequences would be needed for convergence

(if convergence could be achieved at all). Almost all behavioral parameters also have credible 95 percent intervals for sample size $T_1 = 200$, that are essentially identical to those of their priors (a uniform distribution over the interval $[0,1]$). The only exception is parameter b_1 for which the credible 95 percent interval is concentrated below its true value. The Markov chains in the bottom panel of Figs. 1 through 3 show that each chain is wandering around unsystematically without any apparent attraction towards the “true” value. It is worth while to note that the reason for this behavior does not lie in the construction of the chain. On the contrary, the adaptive adjustment of the proposed distribution achieves an acceptance rate that is usually considered optimal in the literature. As we will show below, this aimless movement rather reflects the motion along a likelihood function which is extremely flat along many dimensions. Having one or two orders more data at hand, $T_2 = 2,000$ or $T_3 = 20,000$, improves many estimates gradually. For $T_2 = 2,000$, the improvement is most pronounced for parameters a_2 and b_2 , weaker for a_1 and b_1 , and absent for ρ and γ . For parameters a_1 and b_1 , the different initialisations of the chain still lead to distinctly different dynamic evolutions although they are now concentrated in a relatively narrow interval around the true value. With large enough data $T_3 = 20,000$, the chain appears to converge to a narrow interval for the four slope parameters of the IS and Philipps curve. For parameters ρ and γ , no progress at all is observed even for the largest sample size. This signals the clear danger that any naive estimation

of one of these parameters of the model might just bring about any one of a number of possible inconsistent estimates. The convergence diagnostic indicates this problem by highlighting the lack of convergence even with very long samples, as none of the behavioral parameters features a scale reduction factor that is anywhere close to its target (of about 1.1)

The three variances of the random innovations appear well behaved with nicely converging posteriors. Their \hat{R} statistics are hovering between 1 and 2, thus indicating some tendency of convergence, but scope for improvements with still larger sequences.

While the chains of the innovation variances (not shown in the figures) typically show less variability than those of the behavioral parameters even for small sample sizes, they are nevertheless not identical across the 10 sequences: The intervals in which they move appear slightly different, depending on the tendencies of the other parameters which explains their not-fully-satisfactory convergence statistics.

Hence, practically all structural parameters are not identified in this model at least for typical macroeconomic sample sizes of the order of 10^2 observations. For the new (behavioral) parameters ρ and λ , not even a sample size of order 10^4 provides for any improvement against our uninformative prior. One reason for this behavior is a very flat gradient of the likelihood along many directions. This can even be seen for a seemingly innocent parameter like σ_1 for which we track the development of partial likelihoods over different sample sizes, here extending the sample size even to

$T = 200,000$, in Fig. 4. As can be seen, the partial likelihood is almost flat over a large range of values around the ‘true’ one. Fig. 5 shows a similar clear tendency for a bivariate plot of the partial likelihood for γ and ρ (with the differences in likelihood so small across a wide range of parameter values even for 200,000 data points that the global maximum is hard to spot). Fig. 6 shows a somewhat different behavior in a bivariate plot of parameters a_1 and a_2 with a ridge at the true value $a_2 = -0.5$ together with a low degree of sensitivity with respect to the other parameter, a_1 .

What are the deeper structural reasons for the problems to identify the underlying parameters of the behavioral NKM model? To get some clue to the structural properties that impede recovery of the ‘true’ parameters, we express the dynamics of the observable quantities y_t , π_t and r_t in the form of a system of three difference equations with time-varying parameters $P_{e,t}^g$ and $P_{e,t}^\pi$ for the weight of extrapolative expectations that summarizes the effect of the latent behavioral part of eqs. (2) and (3):

$$\begin{pmatrix} g_t \\ \pi_t \\ r_t \end{pmatrix} = \frac{1}{\Delta} \begin{pmatrix} 1 - a_1 + a_1 P_{e,t}^g & a_2 c_1 (1 - b_1 + b_1 P_{e,t}^\pi) - a_2 P_{e,t}^\pi & a_2 c_3 \\ b_2 (1 - a_1 + a_1 P_{e,t}^g) & (1 - a_2 c_2) (b_1 P_{e,t}^\pi + 1 - b_1) - a_2 b_2 P_{e,t}^\pi & a_2 b_2 c_3 \\ (b_2 c_1 + c_2) (a_1 P_{e,t}^g + 1 - a_1) & c_1 (b_1 P_{e,t}^\pi + 1 - b_1) - (b_2 c_1 + c_2) P_{e,t}^\pi & c_3 \end{pmatrix} \begin{pmatrix} g_{t-1} \\ \pi_{t-1} \\ r_{t-1} \end{pmatrix} \\
+ \frac{1}{\Delta} \begin{pmatrix} \sigma_1 & a_2 c_1 \sigma_2 & a_2 \sigma_3 \\ b_2 \sigma_1 & (1 - a_2 c_2) \sigma_2 & a_2 b_2 \sigma_3 \\ (b_2 c_1 + c_2) \sigma_1 & c_1 \sigma_2 & \sigma_3 \end{pmatrix} \begin{pmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{pmatrix}, \quad (7)$$

with $\Delta = 1 - a_2(b_2 c_1 + c_2)$. The system of eqs. (7) is obtained after eliminating the contemporaneous effects (coming with parameters a_2 , b_2 , c_1 and c_2) on the right-hand side of eqs. (1). In the second part, we express the stochastic terms in the form of three random variates drawn independently from the standard Gaussian with variance of unity. One may note that the presence of both contemporaneous and lagged endogenous variables in eqs. (1) could already constitute an impediment to full identification of all parameters. In a dynamic model with a full matrix of both contemporaneous and lagged effects, the reduced form analogue to eqs. (7) would only contain half as many effects as there are parameters to estimate. Hence, an infinite set of parameter values would be observationally equivalent. The rich literature on

structural VAR models is to a large part concerned with finding auxiliary conditions that allow identification of such a priori over-parametrized models (cf. Baumeister and Hamilton, 2021). In linearized DSGE models, the structured parameters are functions of the ‘deep’ economic parameters, and a necessary condition for identification is that the number of dynamic effects in the VAR representation is not smaller than the number of deep parameters one wishes to estimate. While our model is nonlinear, in a linearized version this condition would be met, since we have 9 dynamic effects on the right hand side of eq. (7) for six parameters to be estimated from this structure $(a_1, a_2, b_1, b_2, \gamma \text{ and } \rho)$.

The resulting system is indeed relatively close to a linear system since both $P_{e,t}^g$ and $P_{e,t}^\pi$ fluctuate around 0.5 as in the equilibrium, $g^* = \pi^* = r^* = 0$, so that none of both forecast heuristics has an advantage over the alternative. What we can see is that just as in the case of certain DSGE models (Canova and Sala, 2009), the dynamic coefficients of the system of difference equations are complex functions of the structural parameters. Considering the variance-covariance matrix of the shock terms one notes that it has non-zero entries throughout which consist also of more or less complex functions of the ‘deep’ parameters. This means that the observed variability of each observed variable depends on the shocks to all three endogenous variables. Considering, for instance, the variance of output growth, we would maintain the same variance with decreasing σ_1 and similarly decreasing a_2 (in absolute value, since a_2 is

negative by definition).

Other complex interactions between structural parameters could lead to similar compensating changes. The very flat shape of the objective function for many parameters makes it easy to compensate for the higher/lower values of some parameters by appropriate changes of others. This problem is exacerbated by a low sensitivity of the objective function with respect to individual parameters even in the case where the remaining ones assume their ‘true’ values. Fig. 4 shows, for instance, that the partial likelihood shows very little sensitivity for all values of σ_1 from half the ‘true’ one to at least twice this value. This flatness explains why large data sets are needed to obtain a posterior distribution that differs significantly from the prior.

If $P_{e,t}^g$ and $P_{e,t}^\pi$ were linear functions of the parameters γ and ρ , eq. (7) would constitute a linear VAR system of first order. Its nine dynamic effects would, then, constitute the information to extract the six behavioral parameters, a_1 , a_2 , b_1 , b_2 , γ and ρ , and the variance part would form a system of three equations to retrieve σ_1 , σ_2 and σ_3 . Even then, however, the nonlinearity and multiplicative form in which the behavioral parameters enter the structural coefficients might make identification cumbersome. For instance, the first and third entries of the second equation are equal to their counterparts in the first equation multiplied by two. Hence, at least for certain sets of parameters, the dynamic coefficients for g_t and π_t are close to colinear. The same actually applies to the variance-covariance matrix. The nonlinear

effects contributed by parameters γ and ρ might further confound these obstacles to identification of the behavioral parameters.

Lack of identification does, however, not necessarily prevent the use of the model for inference on the systematic part of the dynamic evolution. Viewing the shocks as noise that conceals the systematic motion of y_t , π_t and r_t , we can use the estimated parameters for filtering the conditional expectations. In Fig. 7, we use the posterior means of one of the 10 Markov chains for the case $T_1 = 200$, for filtering conditional expectations comparing the results with alternative ones obtained on the base of the true parameters. As Fig. 7 shows, filtering with the mostly biased estimates leads to almost the same result as filtering using the ‘true’ parameter values. As it turns out, the filtered trajectories are virtually identical for both sets of parameters for the inflation and growth rates. We can also see in the right-hand panels of the figure that the filtered fractions of fundamentalists are quite close to the true ones (because of the absence of noise in the state equation, with the ‘true’ parameter values one retrieves the ‘true’ fractions of the heuristics). In the example shown here there is often a slightly higher fraction of fundamentalists estimated with the ‘wrong’ set of parameter values, but the filtered fluctuations of the relative fractions of both heuristics are nicely aligned with those of the underlying sample.

4 An Alternative Specification of the Behavioral NKM Model

While only one specification of de Grauwe’s behavioral NKM model has been studied above, we have encountered similar problems of identifiability of parameters in other versions as well, using alternative formalizations of the forecast heuristics. This begs the question whether we have to expect lack of identification to be an intrinsic defect of behavioral macroeconomic models. It actually turns out that this is not necessarily the case. A slightly different version of model (1) indeed leads to apparent identification with nice convergence properties of all parameters (including ρ and γ), still using the behavioral components specified in eqs. (2) and (3).³

The only change we make to the model is to base the Taylor rule on expected quantities, rather than contemporaneous deviations from target. The modification follows Clarida et al. (2000) and appears even closer to the spirit of the baseline ‘rational’ NKM approach:

$$r_t = c_1(E\pi_{t+1} - \pi^*) + c_2Eg_{t+1} + c_3r_{t-1} + \epsilon_{3,t} \quad (8)$$

³Kukačka and Sacht (2023) also develop a version of the model in which all parameters are identified. However, they arrive at identification by fixing a large number of the parameters of the original model and so, probably, remove most of the interference between parameters visible in eq. (7).

Note that this version of the behavioral NKM assumes that the central bank forms its expectations in the same way like the general public, i.e. using the discrete choice formalization of eqs. (2) and (3). In a recent paper, de Grauwe and Ji (2023) compare the efficiency of monetary policy between a behavioral NKM with forward-looking Talor rule as in eq. (8) and the version based on current macroeconomic data of the previous section. They find that eq. (8) is less effective in stabilizing macroeconomic outcomes. One might not expect a priori that this version is easier to estimate since, on the one hand, the change to expectations now makes the nonlinear behavioral dynamics also enter into the determination of r_t (which had been linear before), and, on the other hand, the parameters assigned to these components, c_1 and c_2 , are not entering anyway in the estimation exercise.

However, as Table 2 and Fig. 8 show, all posteriors are converging in this case to their ‘true’ underlying values with a high degree of precision already obtained for intermediate sample sizes $T_2 = 2000$. This experiment exemplifies that identification of parameters is, in principle, possible with highly nonlinear behavioral models. Fig. 8 depicts the development of the Markov chains exemplarily for parameters γ and ρ . Note that in this case, also the notoriously difficult parameters of the discrete choice part are well identified although for the three-digit sample sizes common in macroeconomics there is still not much improvement away from the prior. However, for longer samples the precision of the posteriors of γ and ρ is about the same as that

of the linear NKM parameters in this setting. The key to identification is probably the simpler format of the reduced form of this system:

$$\begin{pmatrix} g_t \\ \pi_t \\ r_t \end{pmatrix} = \begin{pmatrix} 1 - a_1 + a_1 P_{e,t}^g + a_2 c_2 P_{e,t}^g & -a_2(1 - c_1)P_{e,t}^\pi & a_2 c_3 \\ b_2(1 - a_1 + a_1 P_{e,t}^g + a_2 c_2 P_{e,t}^g) & -a_2 b_2 P_{e,t}^\pi + b_1 P_{e,t}^\pi + 1 - b_1 + a_2 b_2 c_1 P_{e,t}^\pi & a_2 b_2 c_3 \\ c_2 P_{e,t}^g & c_1 P_{e,t}^\pi & c_3 \end{pmatrix} \begin{pmatrix} g_{t-1} \\ \pi_{t-1} \\ r_{t-1} \end{pmatrix} + \begin{pmatrix} \sigma_1 & 0 & a_2 \sigma_3 \\ b_2 \sigma_1 & \sigma_2 & a_2 b_2 \sigma_3 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{pmatrix} \quad (9)$$

In eq. (9), we see that, in particular, the variance-covariance matrix features less involved expressions of the structural parameters. In this system, identification of the shock variables could be done sequentially: First, σ_3 is trivially obtained as the standard deviation of conditional expectations of interest rates. With σ_3 determined in this way, the first row provides σ_1 (if a_2 were known) and the second row finally identifies σ_2 (if b_2 were obtained from the dynamic coefficients). The relative ease of identifying ρ and λ might stem from the fact that the dynamic coefficients for r_t with respect to g_{t-1} and π_{t-1} only depend on the Taylor reaction coefficients, multiplied by the current fraction of extrapolators, which is mainly governed by parameters ρ

and γ .⁴

5 An Empirical Application

We proceed with a heroic effort to estimate the parameters of the modified behavioral NKM model of the previous section with empirical data. We use data on the output gap, inflation rate based on the consumer price index and the federal funds rate over the long time interval from 1954, quarter 3 to 2021, quarter 3, a total of 269 observations.⁵ The data are computed as percentage deviations from their steady state to conform to the framework used in the previous sections. The data transformation follows the recipe of Pfeifer (2021) for the specifications of the data to be used in DSGE models in order to align the observation equations with underlying theory. A slightly shorter subset of these data has been used recently by Kukačka and Sacht (2023) in a related study of a version of the de Grauwe’s behavioral NKM framework. As before, we fix the parameters of the Taylor rule and attempt to estimate the remaining nine parameters. A simple regression indicates a very high degree of auto-correlation in

⁴Iskrev (2010) shows the following necessary condition for local identification of the deep parameters of linearized DSGE models: The Jacobian of the stacked elements of the matrices of the reduced form (i.e., a format similar to the ones in eqs. (7) and (9)) has to have full column rank. The Jacobian in this case is defined as the matrix of the derivatives of all the entries of the matrix of structural parameters that are not independent of the ‘deep’ economic parameters with respect to the latter. If we linearize the time-varying components of the structural matrices of eqs. (9) assuming that $P_{e,t}^g$ and $P_{e,t}^\pi$ are both functions of γ and ρ , and that the linear approximations of the functions are not identical, we indeed can easily demonstrate that the so-defined Jacobian has full column rank (i.e. rank equal to 9, the number of parameters to be estimated). For eqs. (7), the rank of the Jacobian is less obvious.

⁵I am grateful to Stephen Sacht for providing me with the transformed data.

the interest rate series of 0.97. In contrast, the correlation coefficients for the output gap and for inflation are substantially below 0.1 both. The usual Taylor rule, thus, has been adopted in its usual sluggish form with reaction coefficients multiplied by $1-c_3$. The two coefficients, thus, amount to $c_1 = 1.5 * 0.03 = 0.045$, and $c_2 = 0.5 * 0.03 = 0.015$ in both cases. To estimate the posterior we used the same uniform priors as in the Monte Carlo experiments, again together with 10 independent sequences of adaptive MCMC with a total of 2,001,500 draws in each. Table 3 shows the results based on the last 1,000,000 draws of those 10 chains.

As can be seen from Table 3, adaptation works equally well for the empirical series as with the artificial ones. Convergence of the chain appears fully satisfactory under the Gelman/Rubin criterion for parameters b_1 , b_2 and σ_3 and nearly so for a_2 , σ_1 and σ_2 . The convergence indeed appears even better than for any of our previous Monte Carlo experiments. However, the two crucial behavioral parameters, γ and ρ , are characterized again by credible 95 percent intervals that are not too different from those of their uniform priors so that the data does not seem to provide much information on these parameters. We also estimated a VAR(1) model for comparison (with c_1 , c_2 and c_3 fixed as for the behavioral NKM) which seemed to fit the data better than the behavioral model. Since the behavioral components in the NKM appear to be without much relevance for the fit of the estimation, the dominance of the VAR model is not surprising as the NKM without its nonlinear behavioral

components would essentially be a restricted VAR model.

Comparing in-sample root mean-squared errors of the one-period-ahead predictions of the variable, the behavioral NKM falls also behind the VAR(1) for the output gap, but it predicts inflation marginally better. As Fig. 9 shows, the differences are, indeed, tiny, and both competitors track the empirical data relatively well. Also shown in this graph are the predicted fractions of agents using fundamentalist forecasts for g and π which for both macroeconomic variables show wide swings back and forth. How informative these results are is, however, questionable as the two parameters ruling the choice of heuristics, γ and ρ , are apparently not very precisely identified by our estimation.

6 Conclusion

The present paper has highlighted that behavioral NKM models might suffer from a lack of identifiability of their ‘deep’ parameters, at least for sample sizes common in macroeconomics. These findings explain why a recent stream of literature that attempts to estimate more or less identical versions of a specification proposed by de Grauwe (2012) appear by and large inconclusive and without any obvious overall tendency concerning the behavioral parameters that are retrieved from various empirical applications of this framework. As it turns out, the particular version of the model used in many papers suffer from a very flat likelihood, that prevents, in par-

ticular, the identification of the new behavioral parameters, even if an unrealistically large number of data were available. Since the present approach has been based on the exact likelihood of the data, it seems unlikely that alternative estimations used in the literature (that all used some form of approximation to the likelihood) would perform better. Of course, one could resort to ‘strong’ priors as much of the empirical DSGE literature does. However, as the results for small samples show, this would basically amount to the admission that the data itself does not contribute anything to the identification. Such an approach could easily become completely circular when authors embark on using priors concentrated at estimates of previous papers which themselves would have been the outcome of an estimation using uninformative data.

Such lack of identification is not a new phenomenon in the applied macroeconomic literature: lack of identification has been observed in various specifications of DSGE models (e.g. Canova and Sala, 2009) and adding appropriate conditions for identification of a-priori unidentified model structures is the overarching topic of much of the recent literature on structural VAR models.

In behavioral NKM models, the issue of identifiability has been basically ignored. The present paper shows that such an omission could lead to serious flaws of any attempt at empirical validation of such models that essentially render their results meaningless. Like in other areas, identifiability should, therefore, be addressed before one proceeds to estimation proper.

As an alternative specification of Sec. 4 shows, it depends on the exact details of the model set-up (just like in DSGE models) whether a certain parameter of a behavioral NKM are identified or not. Interestingly, at least with enough data, in the forward-looking version of the model even the ‘very deep’ parameters of the discrete choice component for the selection of forecast heuristics can be retrieved, with an efficiency of the estimates that does not differ too much from that of the traditional linear parameters of the three-equation model. However, even for such a well-behaved specification, the scarcity of macro data (with three-digit samples used for estimation of ten or more parameters) constitutes a serious limitation in practice.

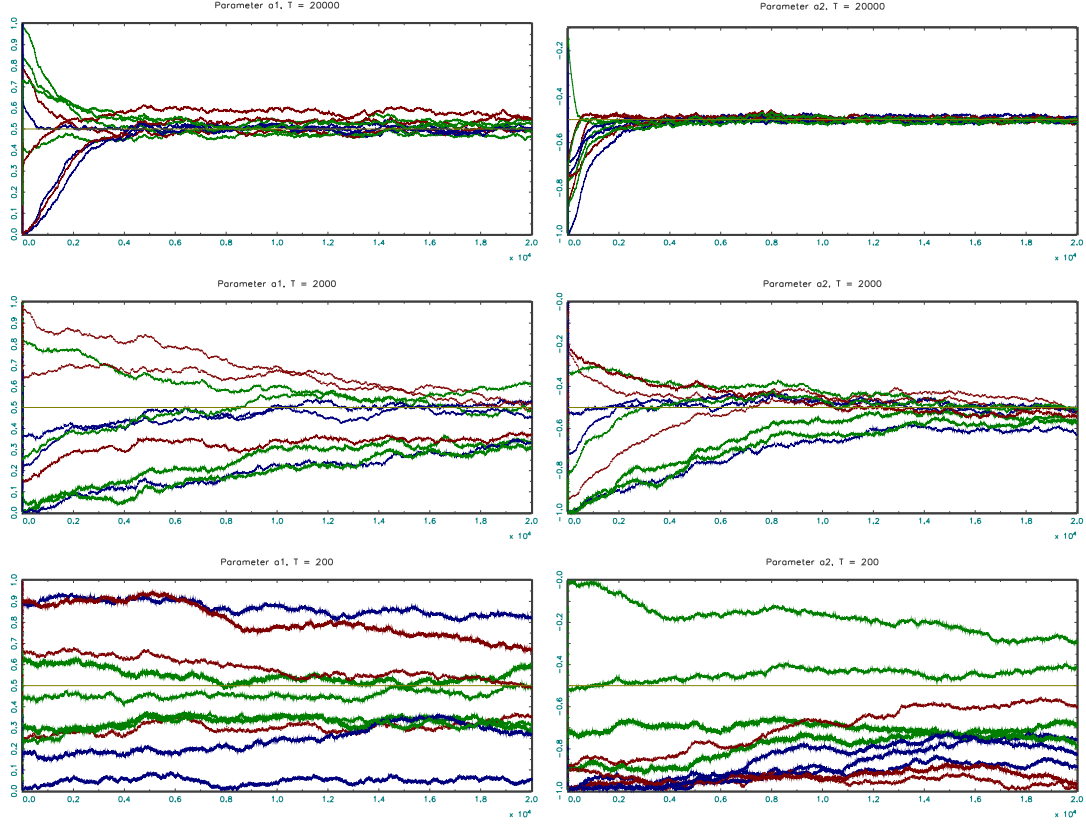


Figure 1: Markov chains for parameters a_1 and a_2 for the model defined by eqs. (1) to (3) from simulations with length $T_1 = 200$, $T_2 = 2,000$ and $T_3 = 20,000$. The lengths of the MCMC experiments is 2,001,500 in all cases, but only every 100th entry is displayed. The straight solid lines depict the ‘true’ parameters of the underlying simulated samples.

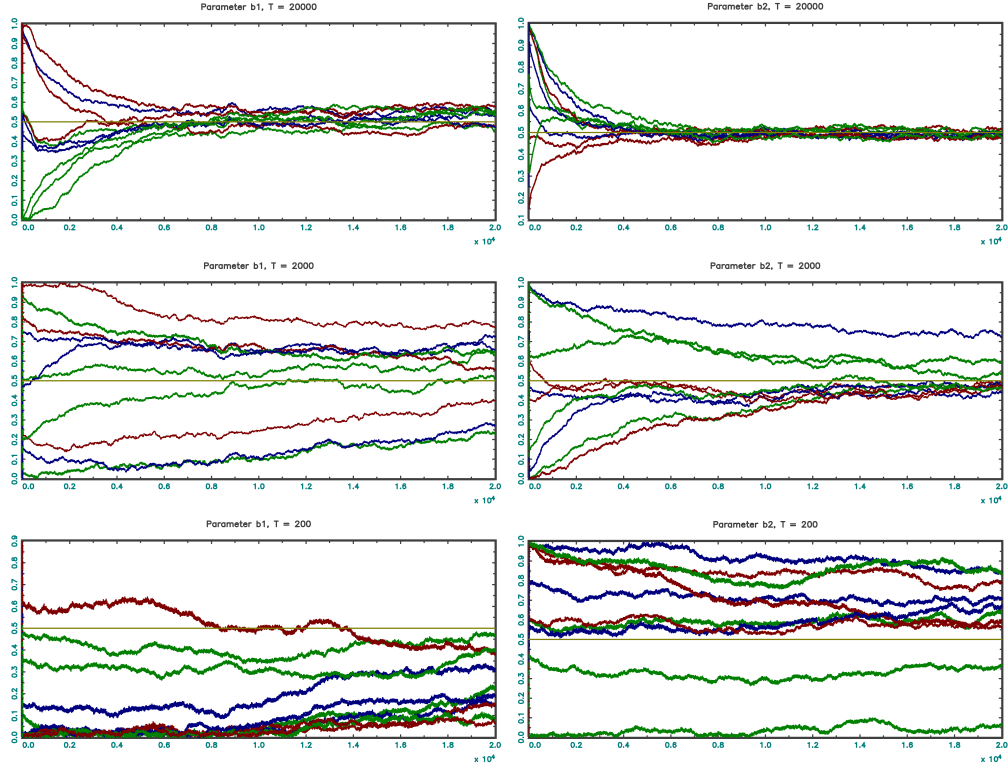


Figure 2: Markov chains for parameters b_1 and b_2 from the same MCMC runs as in Fig. 1.

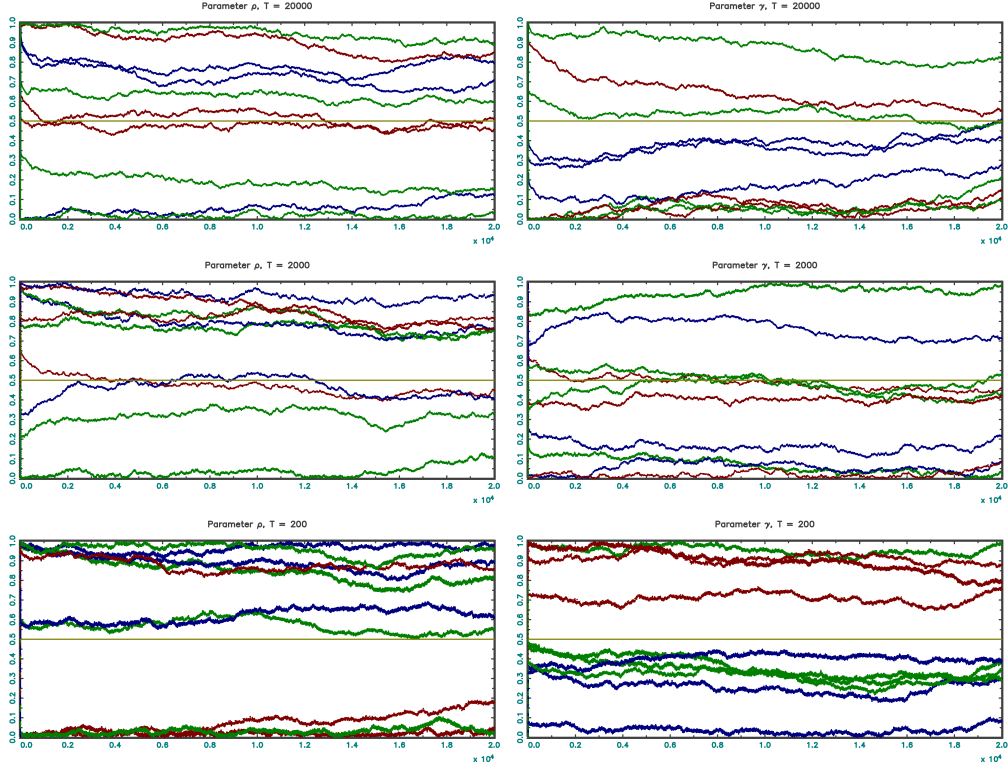


Figure 3: Markov chains for parameters γ and ρ for the same MCMC runs as in Figs. 1 and 2.

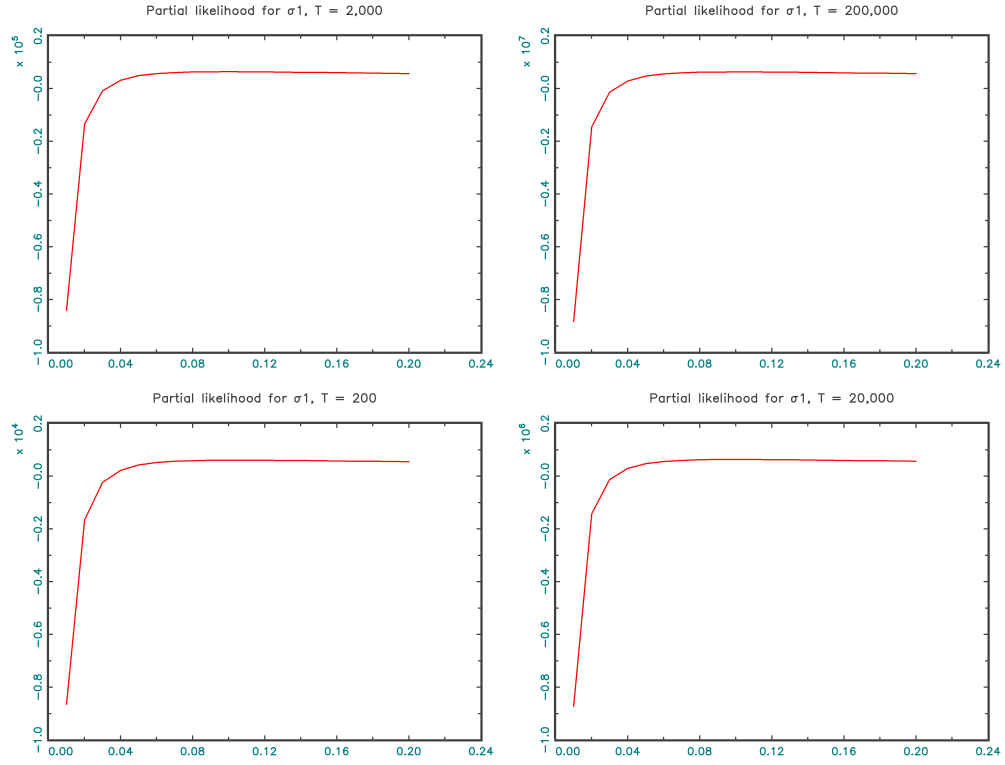


Figure 4: Partial likelihood for parameter σ_1 for different sample sizes from $T = 200$ to $T = 200,000$.

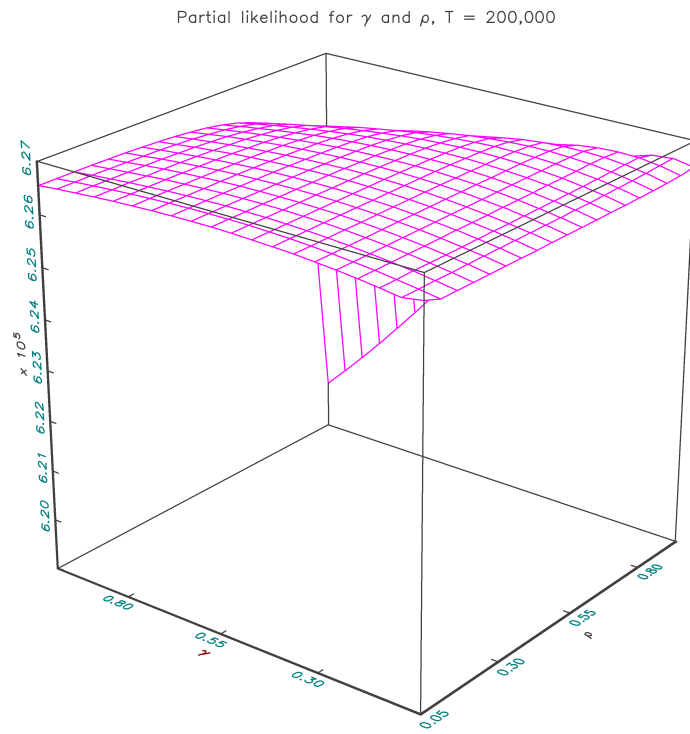


Figure 5: Partial likelihood for parameters γ and ρ for a sample with size $T = 200,000$.

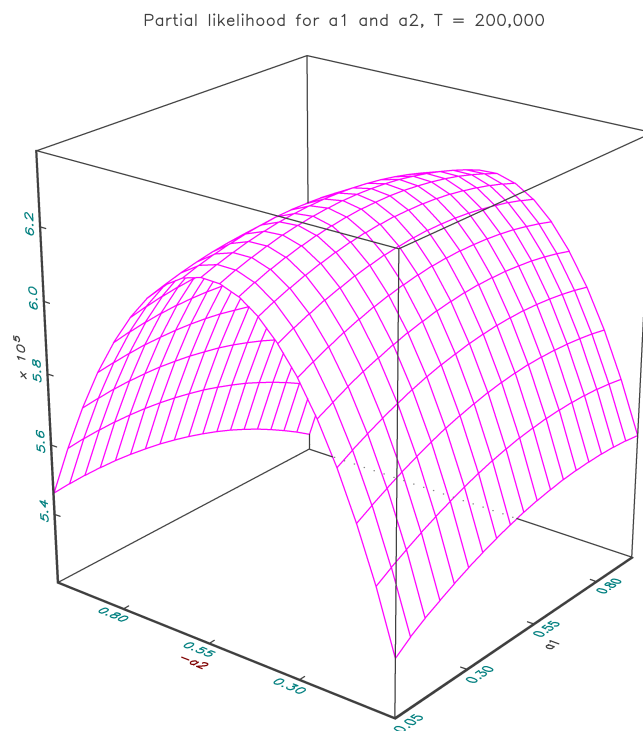


Figure 6: Partial likelihood for parameters a_1 and a_2 for a sample with size $T = 200,000$.

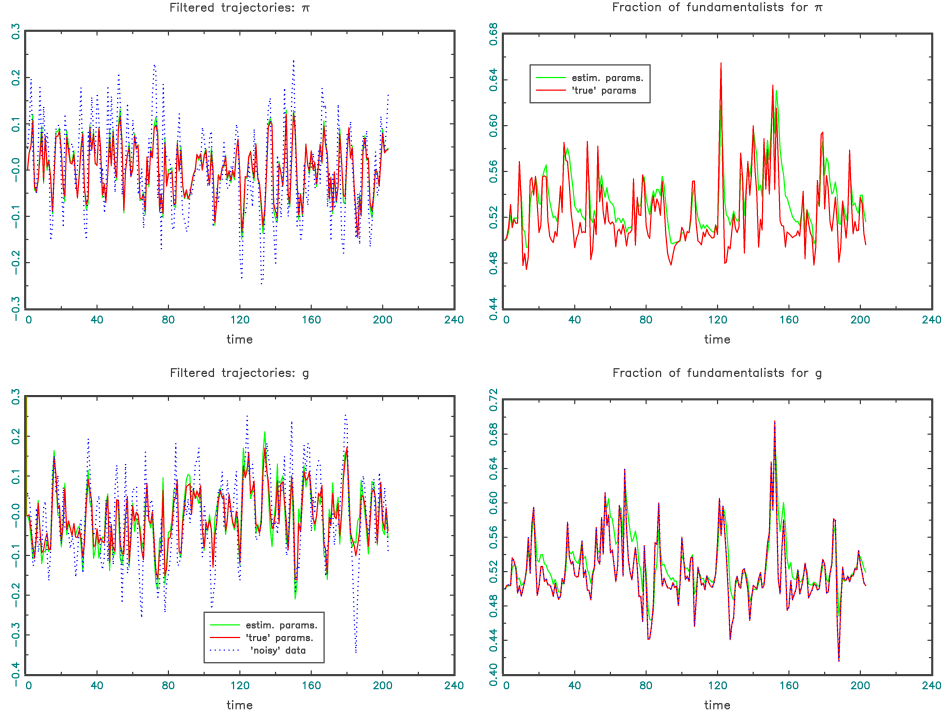


Figure 7: Left-hand side: Filtered trajectories for output gap and inflation for both estimated parameters (means of the posterior of one MCMC run for a sample of size $T_1 = 200$) and ‘true’ parameter values. The estimated parameters were: $a_1 = 0.531$, $a_2 = -0.728$, $b_1 = 0.117$, $b_2 = 0.601$, $\sigma_1 = 0.125$, $\sigma_2 = 0.107$, $\sigma_3 = 0.100$, $\rho = 0.801$, and $\gamma = 0.306$.

Right-hand side: Filtered trajectories of one of the latent variables, the fraction of fundamentalists in the heuristic predictions of the output gap, g , and the inflation rate, π . Note that because of absence of stochastic terms in the equations governing the transitions of latent states, using the ‘true’ parameters enables one to retrieve exactly the time development of these variables.

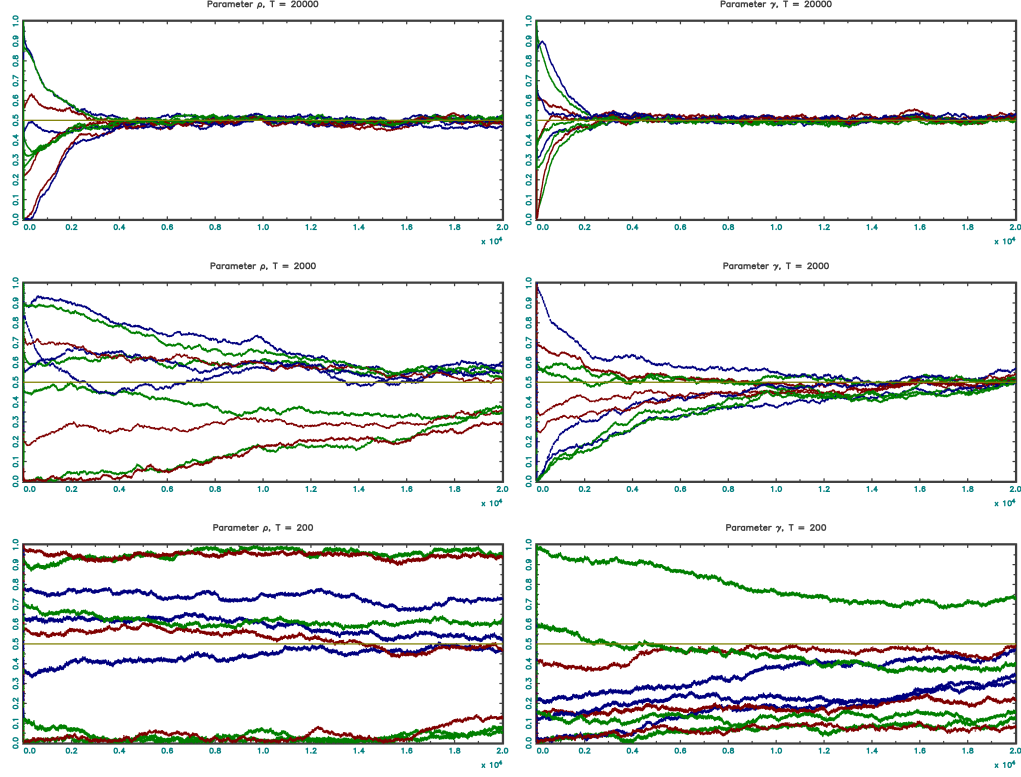


Figure 8: Markov chains for parameters γ and ρ for the model defined by eq. (8) together with the remaining components of the behavioral NKM of de Grauwe (2012) with simulation lengths $T_1 = 200$, $T_2 = 2000$, and $T_3 = 20,000$. The lengths of the MCMC experiments is 2,001,500 in all cases, but only every 100th entry is displayed. The solid straight lines depict the ‘true’ parameters of the underlying simulated samples.

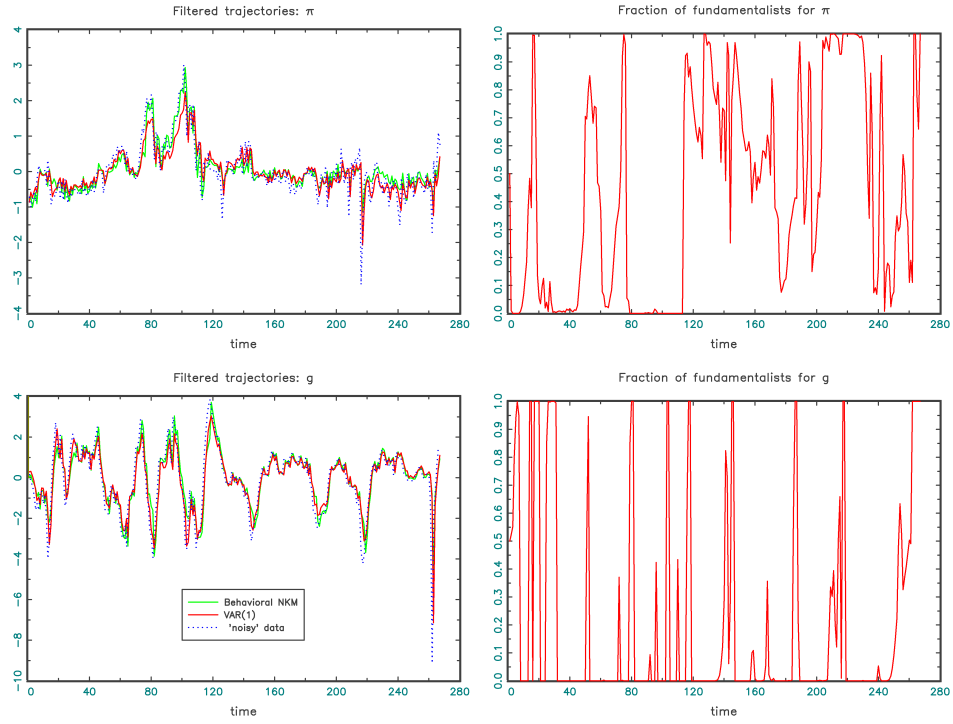


Figure 9:

Left-hand side: Filtered trajectories (i.e., conditional expectations) of the observed variables g and π for the estimated model (means of the posterior distribution shown in Table 3) for U.S. quarterly data.

Right-hand side: Estimated fractions of fundamentalists in the heuristic prediction of output gap, y , and inflation, π .

Table 1: Posterior distribution of parameters

		mean	95% credible int.		\hat{R}	R_{95}
a_1	T1	0.444	0.045	0.857	11.156	42.924
	T2	0.445	0.222	0.659	4.794	19.108
	T3	0.512	0.464	0.589	2.979	5.763
a_2	T1	-0.706	-0.969	-0.194	8.441	33.634
	T2	-0.515	-0.629	-0.432	3.143	7.067
	T3	-0.501	-0.518	-0.481	2.044	2.852
b_1	T1	0.200	0.024	0.499	4.655	8.657
	T2	0.512	0.128	0.800	8.395	25.066
	T3	0.522	0.442	0.583	2.718	5.888
b_2	T1	0.606	0.035	0.907	10.086	29.983
	T2	0.506	0.389	0.777	5.310	15.956
	T3	0.495	0.468	0.528	2.009	3.139
σ_1	T1	0.125	0.085	0.160	2.958	7.338
	T2	0.102	0.095	0.115	2.618	4.605
	T3	0.100	0.098	0.102	1.798	2.388
σ_2	T1	0.109	0.088	0.130	1.994	3.022
	T2	0.101	0.093	0.115	3.320	5.121
	T3	0.100	0.098	0.102	1.833	2.497
σ_3	T1	0.102	0.087	0.121	1.550	2.128
	T2	0.101	0.097	0.105	1.195	1.376
	T3	0.100	0.098	0.101	1.294	1.548
ρ	T1	0.581	0.012	0.985	18.060	54.387
	T2	0.606	0.014	0.926	9.338	22.731
	T3	0.506	0.007	0.936	13.943	40.015
γ	T1	0.498	0.017	0.947	14.639	42.224
	T2	0.371	0.006	0.979	16.610	69.094
	T3	0.317	0.025	0.845	9.332	34.237
logl	T1	603.469	558.302	635.120	8.489	65.105
	T2	6210.893	6130.096	6325.278	10.009	58.206
	T3	62775.981	62501.214	63161.274	81.371	212.384
Acceptance rate:						
	T_1	0.238				
	T_2	0.237				
	T_3	0.236				

Note: The table shows the parameter estimates and maximized likelihood values for the de Grauwe (2012) model obtained from the posterior distribution. The underlying samples consisted of simulations of 2,001,500 data points each. The table includes the mean parameter estimates and the 95% credible interval of the posterior distribution. Adaptation of the proposal variance was started after 1,500 periods, while the last 1,000,000 observations were used to infer the posterior distribution. \hat{R} is the convergence diagnostic by Gelman and Rubin and R_{95} its 95% upper bound.

Table 2: Posterior distribution of parameters

		mean	95% credible int.		\hat{R}	R_{95}
a_1	T1	0.429	0.185	0.862	7.025	16.323
	T2	0.521	0.463	0.595	1.816	9.954
	T3	0.500	0.486	0.516	1.128	1.274
a_2	T1	-0.451	-0.695	-0.206	4.478	14.881
	T2	-0.500	-0.534	-0.466	1.469	2.632
	T3	-0.499	-0.511	-0.489	1.413	1.774
b_1	T1	0.512	0.065	0.829	7.061	212.309
	T2	0.511	0.444	0.591	2.264	5.008
	T3	0.501	0.483	0.522	1.592	2.326
b_2	T1	0.402	0.146	0.567	4.617	21.892
	T2	0.501	0.472	0.529	1.286	1.611
	T3	0.500	0.490	0.510	1.236	1.473
σ_1	T1	0.106	0.091	0.130	1.683	2.624
	T2	0.100	0.096	0.105	1.362	1.664
	T3	0.100	0.099	0.102	1.569	2.004
σ_2	T1	0.112	0.095	0.133	1.544	2.186
	T2	0.101	0.097	0.106	1.596	2.051
	T3	0.100	0.099	0.102	1.387	1.705
σ_3	T1	0.099	0.085	0.113	1.411	1.800
	T2	0.101	0.097	0.106	1.474	1.852
	T3	0.100	0.099	0.101	1.143	1.279
ρ	T1	0.487	0.009	0.976	17.368	87.211
	T2	0.448	0.175	0.637	5.365	26.125
	T3	0.495	0.463	0.522	1.367	2.257
γ	T1	0.296	0.066	0.730	9.145	81.956
	T2	0.476	0.409	0.535	1.534	2.995
	T3	0.503	0.480	0.530	1.369	1.916
logl	T1	497.596	458.244	524.842	2.863	7.233
	T2	5243.478	5165.910	5390.909	13.029	499.840
	T3	52886.368	52530.816	53192.006	73.453	119.080
Acceptance rate:						
	T_1	0.236				
	T_2	0.237				
	T_3	0.238				

Note: The table shows the parameter estimates and maximized likelihood values for the modified de Grauwe (2012) model obtained from the posterior distribution. The underlying samples consisted of simulations of 2,001,500 data points each. The table includes the mean parameter estimates and the 95% credible interval of the posterior distribution. Adaptation of the proposal variance was started after 1,500 periods, while the last 1,000,000 observations were used to infer the posterior distribution. \hat{R} is the convergence diagnostic by Gelman and Rubin and \hat{R} its 95% upper bound

Table 3: Posterior distribution of parameters for empirical data

	mean	95% credible int.		\hat{R}	R_{95}
a_1	0.431	0.093	0.689	2.652	4.070
a_2	-0.087	-0.239	-0.009	1.124	1.458
b_1	0.687	0.518	0.851	1.037	1.094
b_2	0.064	0.026	0.102	1.003	1.006
σ_1	0.912	0.816	1.036	1.237	1.478
σ_2	0.476	0.430	0.526	1.121	1.245
σ_3	0.204	0.186	0.224	1.008	1.017
γ	0.748	0.039	0.968	3.219	3826.961
ρ	0.413	0.050	0.866	1.492	2.212
logl	-482.666	-494.176	-475.289	2.456	3.585
logl(VAR(1)): -453.795					
Acceptance rates: for MCMC estimation of behavioral MKM: 0.231					
In-sample	BNKM	VAR(1)			
RMSEs:					
g	0.952	0.901			
π	0.467	0.478			
r	0.204	0.199			

Note: The table shows the parameter estimates and maximized likelihood values for the empirical US data under the modified de Grauwe (2012) model obtained from the posterior distribution. The underlying samples consisted of simulations of Markov chains of 2,001,500 data points each. The table includes the mean parameter estimates and the 95% credible interval of the posterior distribution. Adaptation of the proposal variance was started after 1,500 periods, while the last 1,000,000 observations were used to infer the posterior distribution. \hat{R} is the convergence diagnostic by Gelman and Rubin and R_{95} its 95% upper bound. The bottom part of the table compares the root mean squared errors (RMSEs) of conditional expectations of the state variables obtained from the behavioral NKM and a VAR(1) process.

References

- Akerlof, G. A. and R. J. Shiller (2010), *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*, Princeton University Press.
- Andrieu, C. and J. Thomas (2008), A tutorial on adaptive MCMC, *Statistics and Computing* 18 4, 343–373.
- Anufriev, M., T. Assenza, C. Hommes and D. Massaro (2013), Interest rate rules and macroeconomic stability under heterogeneous expectations, *Macroeconomic Dynamics* 17 8, 1574–1604.
- Baumeister, C. and J. D. Hamilton (2021), Advances in using vector autoregressions to estimate structural magnitudes, *Econometric Theory* , 1–39.
- Branch, W. A. and B. McGough (2009), A new Keynesian model with heterogeneous expectations, *Journal of Economic Dynamics and Control* 33 5, 1036–1051.
- Branch, W. A. and B. McGough (2010), Dynamic predictor selection in a new Keynesian model with heterogeneous expectations, *Journal of Economic Dynamics and Control* 34 8, 1492–1508.
- Canova, F. and L. Sala (2009), Back to square one: Identification issues in DSGE models, *Journal of Monetary Economics* 56 4, 431–449.

- Clarida, R., J. Galí and M. Gertler (2000), Monetary policy rules and macroeconomic stability: Evidence and some theory, *Quarterly Journal of Economics* 115 1, 147–180.
- de Grauwe, P. (2011), Animal spirits and monetary policy, *Economic Theory* 47, 423–457.
- de Grauwe, P. (2012), Booms and busts in economic activity: A behavioral explanation, *Journal of Economic Behavior & Organization* 83 3, 484–501.
- de Grauwe, P. and Y. Ji (2020), Structural reforms, animal spirits, and monetary policies, *European Economic Review* 124, 103395.
- de Grauwe, P. and Y. Ji (2023), On the use of current and forward-looking data in monetary policy: a behavioural macroeconomic approach, *Oxford Economic Papers* 75 2, 526–552.
- delli Gatti, D., S. Desiderio, E. Gaffeo, P. Cirillo and M. Gallegati (2011), *Macroeconomics from the bottom-up*, volume 1, Springer Science & Business Media.
- delli Gatti, D. and J. Grazzini (2020), Rising to the challenge: Bayesian estimation and forecasting techniques for macroeconomic agent based models, *Journal of Economic Behavior & Organization* 178, 875–902.
- Douc, R., P. Doukhan and E. Moulines (2013), Ergodicity of observation-driven time

- series models and consistency of the maximum likelihood estimator, *Stochastic Processes and their Applications* 123 7, 2620–2647.
- Douc, R., F. Roueff and T. Sim (2015), Handy sufficient conditions for the convergence of the maximum likelihood estimator in observation-driven models, *Lithuanian Mathematical Journal* 55 3, 367–392.
- Flegal, J. M., M. Haran and G. L. Jones (2008), Markov Chain Monte Carlo: Can we trust the third significant figure?, *Statistical Science* 23 2, 250–260.
- Gelman, A. and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Statistical science* 7 4, 457–472.
- Grazzini, J., M. G. Richiardi and M. Tsionas (2017), Bayesian estimation of agent-based models, *Journal of Economic Dynamics and Control* 77, 26–47.
- Iskrev, N. (2010), Local identification in DSGE models, *Journal of Monetary Economics* 57 2, 189–202.
- 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 (2021), Forecast heuristics, consumer expectations, and New-Keynesian macroeconomics: A horse race, *Journal of Economic Behavior & Organization* 182, 493–511.

- Kukačka, J. and S. Sacht (2023), Estimation of heuristic switching in behavioral macroeconomic models, *Journal of Economic Dynamics and Control* in press.
- Lines, M. and F. Westerhoff (2010), Inflation expectations and macroeconomic dynamics: The case of rational versus extrapolative expectations, *Journal of Economic Dynamics and Control* 34 2, 246–257.
- Lines, M. and F. Westerhoff (2012), Effects of inflation expectations on macroeconomic dynamics: extrapolative versus regressive expectations, *Studies in Nonlinear Dynamics & Econometrics* 16 4, 1–30.
- Lux, T. (2021), Can heterogeneous agent models explain the alleged mispricing of the S&P 500?, *Quantitative Finance* 21 9, 1413–1433.
- Lux, T. (2022), Bayesian estimation of agent-based models via adaptive particle Markov Chain Monte Carlo, *Computational Economics* 60, 451–477.
- McManus, D. A. (1992), How common is identification in parametric models?, *Journal of Econometrics* 53 1-3, 5–23.
- Pfeifer, J. (2021), A guide to specifying observation equations for the estimation of DSGE models, *Manuscript, University of the Federal Armed Forces, Munich* .
- Poledna, S., M. G. Miess, C. Hommes and K. Rabitsch (2023), Economic forecasting with an agent-based model, *European Economic Review* 151, 104–306.

- Rosenthal, J. (2011), Optimal proposal distributions and adaptive MCMC, in S. Brooks, A. Gelman, G. Jones and X.-L. Meng, editors, *Handbook of Markov Chain Monte Carlo*, Cambridge: Chapman & Hall, pages 93–112.
- Siekmann, I., J. Sneyd and E. J. Crampin (2012), MCMC can detect nonidentifiable models, *Biophysical Journal* 103 11, 2275–2286.
- Smets, F. and R. Wouters (2007), Shocks and frictions in US business cycles: A Bayesian DSGE approach, *American Economic Review* 97 3, 586–606.
- Thaler, R. H. (2015), *Misbehaving: The making of behavioral economics*, WW Norton & Company.
- Zhang, J., Q. Zhang, Y. Li and Q. Wang (2023), Sequential Bayesian inference for agent-based models with application to the Chinese business cycle, *Economic Modelling* in press.

Appendix A

Adaption of the Proposal Distribution in Markov Chain Monte Carlo

Satisfactory performance of an MCMC estimation requires a sufficiently high (but also not too high) acceptance rate of the proposals. To define a priori a distribution of the proposals that provides for a reasonable acceptance rate is challenging as acceptances depend on the properties of the unknown posterior distribution. Many applied papers, therefore, include a more or less extended ‘experimental’ part in which different distributions of the proposals are tested before moving to the estimation proper. The statistics literature has developed a more systematic approach towards selection of an appropriate proposal distribution, namely Adaptive Markov Chain Monte Carlo as proposed by Andrieu and Thomas (2008) and Rosenthal (2011). Their proposed algorithms use iterative adjustments of the proposal distribution ‘on the fly’ using the statistical features of those proposals that have already been accepted. One would, then, for instance, update the mean μ_ξ and variances $\hat{\Sigma}_\xi$ of the proposal distribution

as follows:

$$\begin{aligned}\mu_{\xi+1} &= \mu_{\xi} + \gamma_{\xi+1}(\theta_{\xi+1} - \mu_{\xi}) \\ \hat{\Sigma}_{\xi+1} &= \hat{\Sigma}_{\xi} + \gamma_{\xi+1}((\theta_{\xi+1} - \mu_{\xi})(\theta_{\xi+1} - \mu_{\xi})^T - \hat{\Sigma}_{\xi}).\end{aligned}\tag{A1}$$

The degree of adaption, identified by the factor γ_{ξ} , is required to vanish asymptotically to guarantee converge of the chain. In order to optimally scale the variance-covariance matrix, one adds a scale factor λ when drawing new proposals θ^* , using $\lambda\hat{\Sigma}_{\xi}$ instead of $\hat{\Sigma}_{\xi}$ from eqs. A1 which also undergoes adaptation as the chain evolves:

$$\ln(\lambda_{\xi+1}) = \ln(\lambda_{\xi}) + \gamma_{\xi+1}(\alpha_{\xi} - \alpha^*).\tag{A2}$$

This approach allows to target a particular acceptance rate, α^* , which for some toy examples has been found to be equal to 0.234. We also adopt this target here. As can be seen in Tables 1 through 3 this adaptive adjustment works well with the artificial and empirical data. Fig. A1 shows the typical oscillatory convergence of λ_{ξ} that has been observed in all our applications. We also experimented with higher targets such as $\alpha^* = 0.4$. While these could also be achieved, the resulting Markov chain showed no superior performance compared to the baseline case (confirming the conjecture that the acceptance rate should also not be too high).

It remains to report one subtle detail: To avoid that the variances of the proposal distribution get stuck at zero during the transient adjustment, a small fraction of

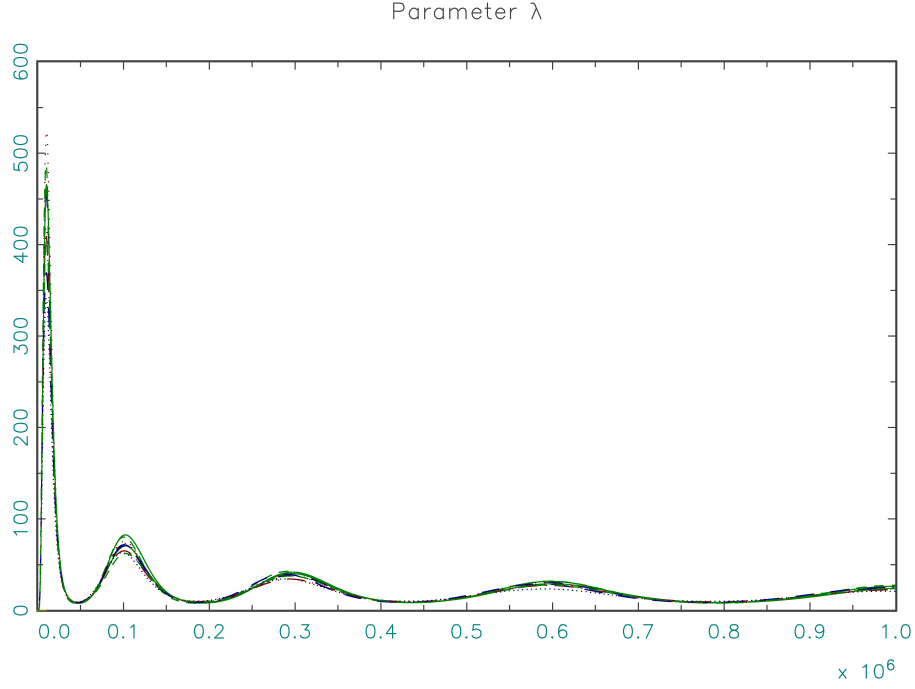


Figure A1: Development of the scale adjustment factors λ of the distribution of proposals for the 10 MCMC runs of the original de Grauwe model formalized by eqs. 1 through 3 with sample size $T_1 = 200$.

the initial variance-covariance matrix has been added, i.e. $\hat{\Sigma}_\xi + \epsilon \hat{\Sigma}_0$, where Σ_0 is the variance that has been used in the initial phase of 1,500 draws without adaption. For Σ_0 , a diagonal matrix with all elements on the main diagonal equal to 0.01 has been used.