**Online Appendices (not for publication)**

Kukacka, J. and Sacht, S. (September 2022), Estimation of heuristic switching in behavioral macroeconomic models.

## Online Appendix A

*Advantages and limitations of the simulated maximum likelihood method*

The SML estimator addresses most of the potential issues related to the alternative estimation approaches mentioned in Section 2. 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 parameterization 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 employ 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 important, 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 because 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 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 is that the standard deviations of the idiosyncratic i.i.d. shocks have to be parameterized because the method requires their distribution to be known.

A promising approach to some extent similar to SML is designed especially for 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, 2021). 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, 2022). However, we are only aware of two rare exceptions applying this Bayesian technique to the estimation of heuristic switching macroeconomic models by Özden (2021) and Fischer (2022).

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.

*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 nonstationary. 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 multidimensional 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 $N \longrightarrow \infty$ and $H \longrightarrow \mathbf{0}$ 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 $\mathbf{0} < H < \bar{H}$ for some $\bar{H}$ as $N \longrightarrow \infty$. 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 facilitates 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 $\hat{p} \longrightarrow p$ sufficiently fast, which even allows for mixed discrete and continuous distributions and the nonstationarity 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

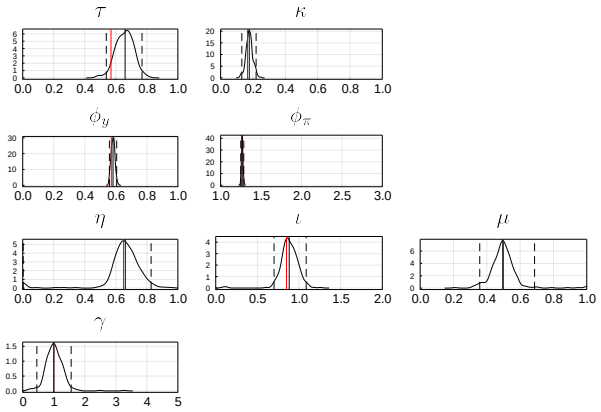
including, e.g., stationarity, the results regarding the higher-order asymptotic properties together with the expressions for the bias and variance components of the SML estimator due to kernel estimation and numerical simulations are derived.

**Online Appendix B: Additional model specifications and robustness analysis**

To support the qualitative robustness of our findings, we repeat the same numerical analysis as in Section 5 for additional alternative realistic parameterizations for both the BR and the hybrid RE NKM. Because of space constraints, 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 parameterizations based on the estimates for the Euro Area data by Jang and Sacht (2021, their Table 3, EFB scenario) and based on the parameterization 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 $\tau$ is also apparent for both parameterizations. Interestingly, for the Jang and Sacht (2021, their Table 3, EFB scenario) parameterization, the bias for $\mu$ disappears, while for the Hommes et al. (2019); De Grauwe and Ji (2020) parameterization, we instead detect new biases for the Taylor rule parameters $\phi_y$ and $\phi_\pi$. The same similarity of the qualitative results holds for the alternative parameterizations of the hybrid RE NKM based on Jang and Sacht (2021, their Tables 2 and 3, REH scenario); De Grauwe and Ji (2020) reported in Figure 4. Interestingly, under the given parameterization, the SML estimator seems to be more accurate for $\alpha$, and the small biases for the structural parameters apparent in Figure 1 seem to diminish. However, $\tau$ 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.

Second, to address the issue of the consistent bias for the inverse intertemporal elasticity of substitution in consumption $\tau$ for the BR NKM, we parameterize its value, and only seven remaining parameters are now subject to estimation. The results for the three alternative parameterizations of $\tau$ are depicted in Figure 5. Although holding this weakly identified coefficient constant does not lead to better estimation performance for the other parameters, the bias does not spill over to the other parameters, and the issue therefore seems to be well resolved.

(a) Jang and Sacht (2021), Table 3, EFB




(b) Hommes et al. (2019); De Grauwe and Ji (2020)

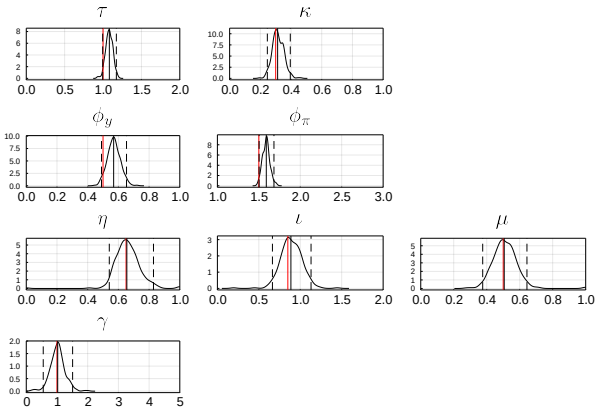
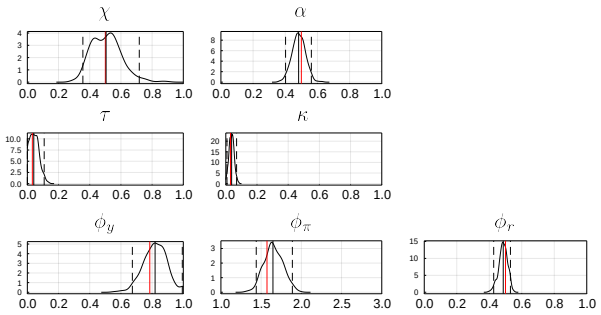



Figure 3: Densities of the pseudo-true parameter estimates for the BR NKM. *Note:* The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, and the dashed red vertical lines depict the 95% confidence intervals of the sample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

(a) Jang and Sacht (2021), Table 2




(b) Jang and Sacht (2021), Table 3

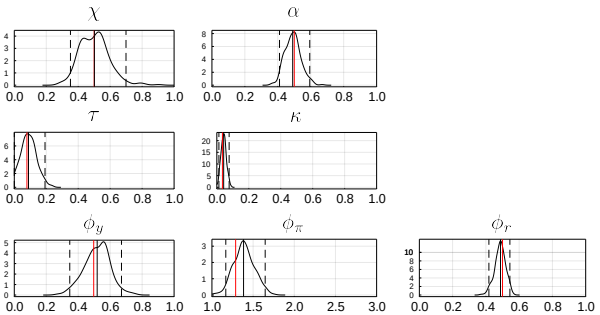
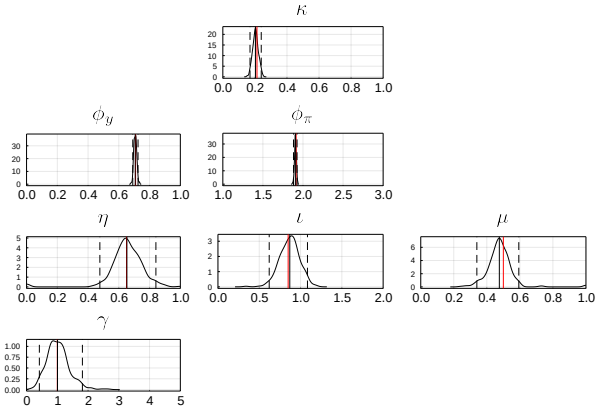
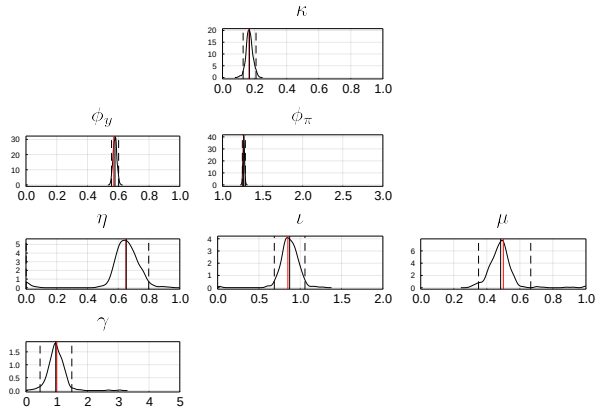



Figure 4: Densities of the pseudo-true parameter estimates for the hybrid RE NKM. *Note:* The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, and the dashed red vertical lines depict the 95% confidence intervals of the sample estimates. Based on 300 random runs, $T = 500$, the parameterization of the intrinsic persistence parameters $\chi$, $\alpha$, $\phi_r$ and the standard deviations of shocks follow De Grauwe and Ji (2020), and the parameterization of the remaining structural parameters is based on Jang and Sacht (2021, their Tables 2 and 3, REH scenarios).

(a) parameterization follows Table 1




(b) Jang and Sacht (2021), Tab. 3, EFB




(c) Hommes et al. (2019); De Grauwe and Ji (2020)

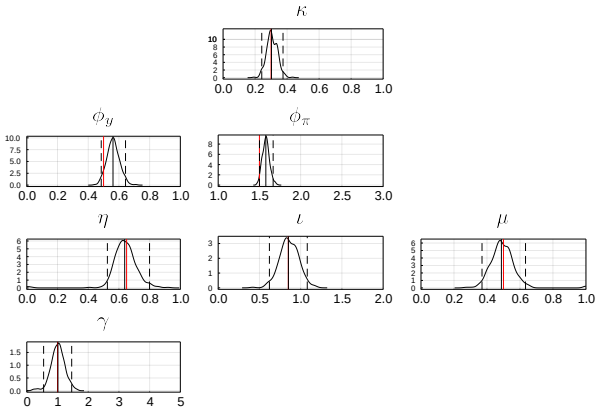



Figure 5: Densities of the pseudo-true parameter estimates for the BR NKM with fixed $\tau$ and alternative parameterizations. *Note:* The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, and the dashed red vertical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

## Online Appendix C: Alternative sets of heuristics

Here 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 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 | $E^F y_{t+1} = \overline{y} + \psi_y(y_{t-1} - \overline{y})$ | Fundamentalists | Gaunersdorfer and Hommes (2007) |
| | $E^C y_{t+1} = y_{t-1} + \xi_y(y_{t-1} - y_{t-2})$ | Chartists | |
| A2 | $E^0_{y_{t+1}} = (1/2)(\beta + \delta\lambda_{y,t})$ | Optimists | De Grauwe (2011) |
| | $E^P_{y_{t+1}} = -(1/2)(\beta + \delta\lambda_{y,t})$ | Pessimists | |
| A3 | $E^f y_{t+1} = \overline{y}$ | 'Strict' fundamentalists | De Grauwe and Ji (2020) |
| | $E^N y_{t+1} = y_{t-1}$ | Naïve ones | |

*Note:* The behavioral parameters $\psi_y, \xi_y, \beta$, and $\delta$ 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 $\overline{y}$.

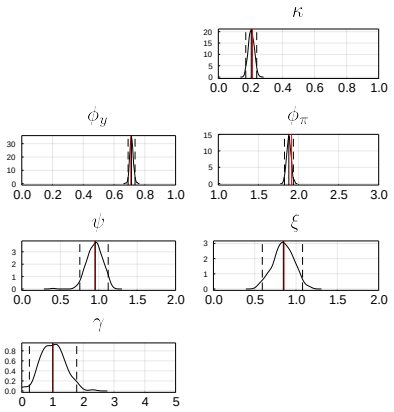
For an in-depth description of alternative sets A1 and A2, we refer directly to Jang and Sacht (2021). With respect to set A3, it becomes apparent that $\psi_y = \xi_y = 0$ holds because we consider the heuristics included in set A1. Note also that in general, we assume zero values for the steady states in output and inflation, i.e., that $\overline{y} = \overline{\pi} = 0$ holds. Inflation expectations are formed consistently for A1-A3 based on two heuristics for the so-called extrapolators ($E$) and targeters ($T$):

$$E^E \pi_{t+1} = \pi_{t-1} \tag{23}$$

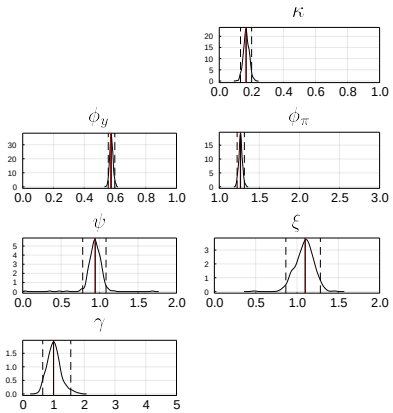
$$E^T \pi_{t+1} = \overline{\pi}. \tag{24}$$

Therefore, regardless of the alternative sets under consideration, inflation expectations are always formed based on the expressions (23) and (24) above. This is in line with the approaches discussed in Jang and Sacht (2021) 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.

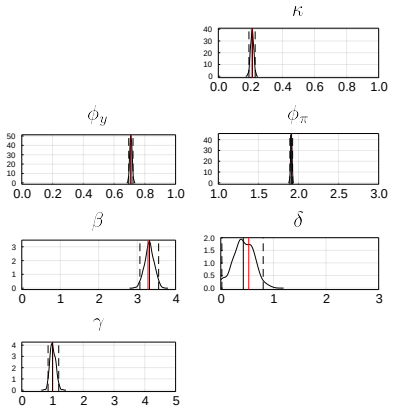
(a) A1: Gaunersdorfer and Hommes (2007)



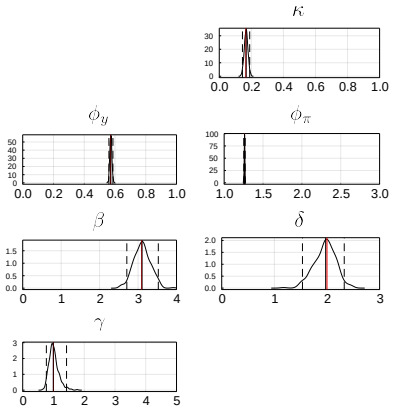

(b) A1: Gaunersdorfer and Hommes (2007)



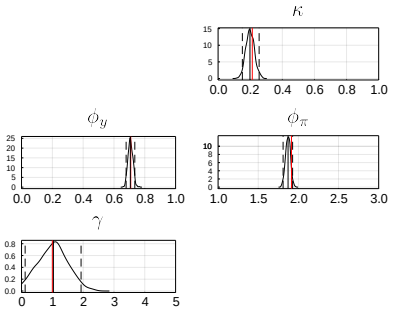

(c) A2: De Grauwe (2011)




(d) A2: De Grauwe (2011)




(e) A3: De Grauwe and Ji (2020)




(f) A3: De Grauwe and Ji (2020)

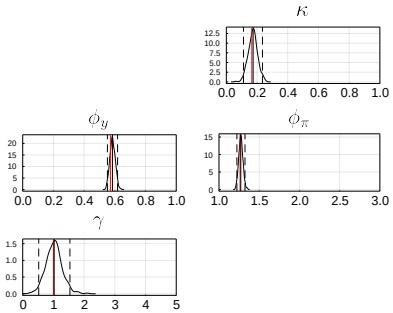



Figure 6: Densities of the pseudo-true parameter estimates for the BR NKM with alternative sets of heuristics. *Note:* The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, and the dashed red vertical lines depict the 95% confidence intervals of the sample estimates. Based on 300 random runs, $T = 500$, the parameterization is based on Jang and Sacht (2021, their Table 2 (left half) and Table 3 (right half), EFB scenarios).

## Online Appendix D: 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 |
| $\tau$ | *.371* | .38 | .37 | .38 |
| $\langle 0,1 \rangle$ | | (.26-.50) | (.29-.47) | (.32-.44) |
| $\kappa$ | *.213* | .21 | .22 | .21 |
| $\langle 0,1 \rangle$ | | (.14-.32) | (.16-.30) | (.17-.29) |
| $\phi_y$ | *.709* | .70 | .71 | .71 |
| $\langle 0,1 \rangle$ | | (.52-.94) | (.56-.90) | (.55-.87) |
| $\phi_\pi$ | *1.914* | 1.94 | 1.95 | 1.97 |
| $\langle 1,3 \rangle$ | | (1.65-2.25) | (1.73-2.17) | (1.77-2.20) |
| $\eta$ | *.65* | .72 | .77 | .62 |
| $\langle 0,1 \rangle$ | | (.02-.99) | (.01-.99) | (.07-.98) |
| $\iota$ | *.85* | .69 | .64 | .70 |
| $\langle 0,2 \rangle$ | | (.13-1.23) | (.16-1.08) | (.28-1.08) |
| $\mu$ | *.50* | .54 | .51 | .54 |
| $\langle 0,1 \rangle$ | | (.23-.79) | (.27-.74) | (.40-.68) |
| $\gamma$ | *1.00* | 1.96 | 2.05 | 2.13 |
| $\langle 0,5 \rangle$ | | (.04-4.72) | (.05-4.70) | (.08-4.64) |

*Note:* 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 parameterization follows Table 1. The figures are rounded to 2 or 3 decimal places. The SMM setup follows Jang and Sacht (2016, 2021); Franke (2022). The reported results are directly comparable to the left half of Table 1.

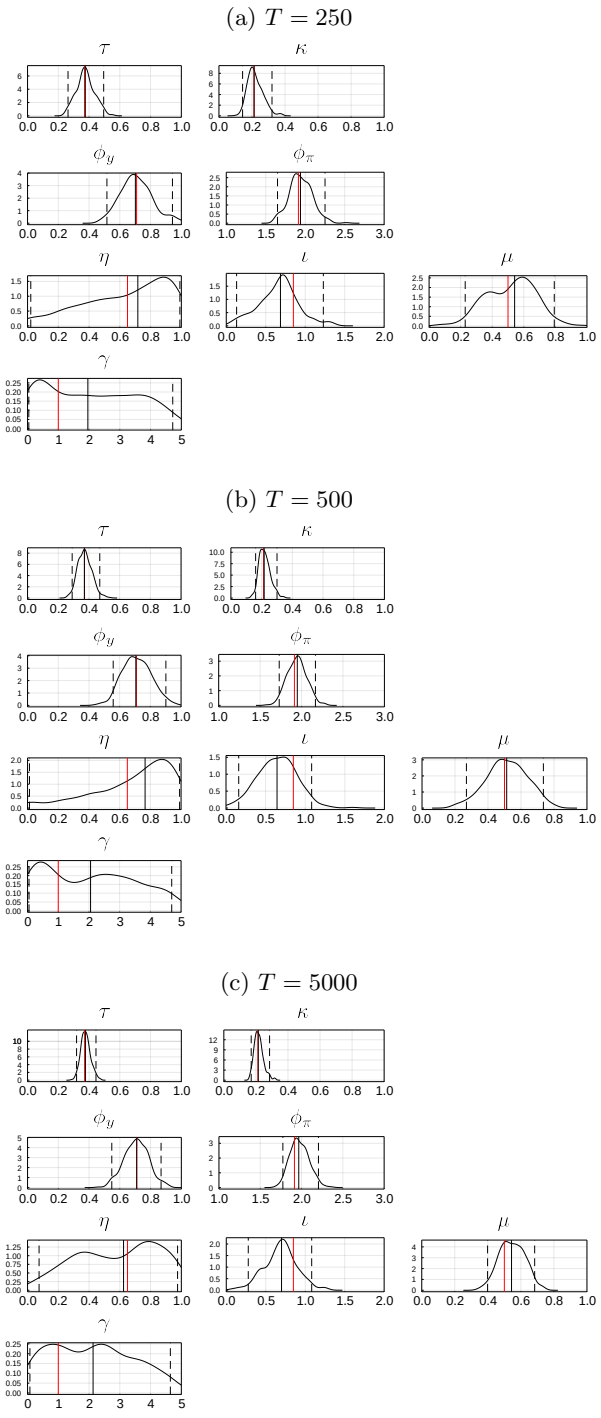



Figure 7: Densities of the pseudo-true parameter estimates for the BR NKM via the SMM following the setup by Jang and Sacht (2016, 2021); Franke (2022). *Note:* The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, and the dashed red vertical lines depict the 95% confidence intervals of the sample estimates. Based on 300 random runs, the parameterization follows Table 1. The reported results are directly comparable to the left half of Figure 1.