Online appendices (not for publication)

Kukacka, J. and Sacht, S. (March 2022), 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 parameteric 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 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 $\hat{p} \longrightarrow p$ 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: 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	*.371*	.38	.37	.38
$\langle 0, 1\rangle$		(.26-.50)	(.29-.47)	(.32-.44)
κ	*.213*	.21	.22	.21
$\langle 0, 1\rangle$		(.14-.32)	(.16-.30)	(.17-.29)
ϕ_y	*.709*	.70	.71	.71
$\langle 0, 1\rangle$		(.52-.94)	(.56-.90)	(.55-.87)
ϕ_π	*1.914*	1.94	1.95	1.97
$\langle 1, 3\rangle$		(1.65-2.25)	(1.73-2.17)	(1.77-2.20)
η	*.65*	.72	.77	.62
$\langle 0, 1\rangle$		(.02-.99)	(.01-.99)	(.07-.98)
ι	*.85*	.69	.64	.70
$\langle 0, 2\rangle$		(.13-1.23)	(.16-1.08)	(.28-1.08)
μ	*.50*	.54	.51	.54
$\langle 0, 1\rangle$		(.23-.79)	(.27-.74)	(.40-.68)
γ	*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 (2019). The reported results are directly comparable to the left half of Table 1.

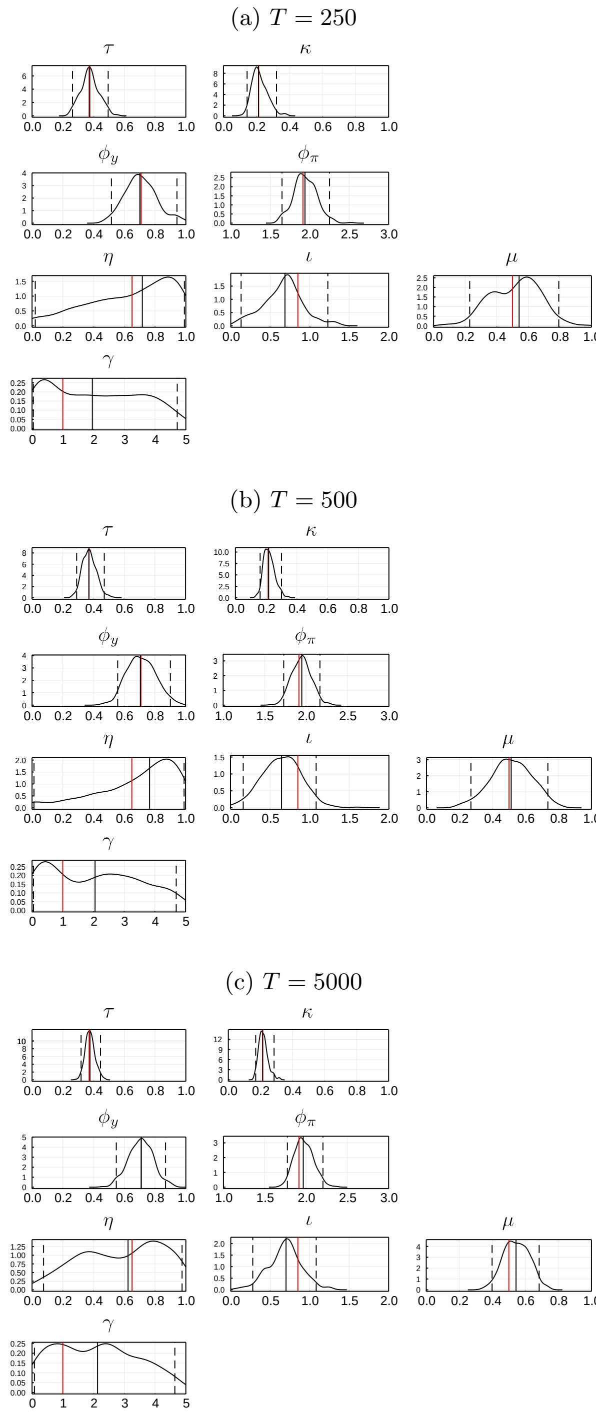

Figure 7: Densities of the pseudo-true parameter estimates for the BR NKM via the SMM following the setup by Jang and Sacht (2016, 2021); Franke (2019). *Note:* The bold black curves depict the kernel density estimates of the sample densities, the bold red vertical lines show the pseudo-true values, 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.