INDIRECT LIKELIHOOD INFERENCE

MICHAEL CREEL AND DENNIS KRISTENSEN

ABSTRACT. Standard indirect Inference (II) estimators take a given finite-dimensional statistic, Z_n, and then estimate the parameters by matching the sample statistic with the model-implied population moment. We here propose a novel estimation method that utilizes all available information contained in the distribution of Z_n, not just its first moment. This is done by computing the likelihood of Z_n, and then estimating the parameters by either maximizing the likelihood or computing the posterior mean for a given prior of the parameters. These are referred to as the maximum indirect likelihood (MIL) and Bayesian Indirect Likelihood (BIL) estimators, respectively. We show that the IL estimators are first-order equivalent to the corresponding moment-based II estimator that employs the optimal weighting matrix. However, due to higher-order features of Z_n, the IL estimators are higher order efficient relative to the standard II estimator. The likelihood of Z_n will in general be unknown and so simulated versions of IL estimators are developed. Monte Carlo results for a structural auction model and a DSGE model show that the proposed estimators indeed have attractive finite sample properties.

Keywords: Approximate Bayesian Computation; Indirect Inference; maximum-likelihood; simulation-based methods.

JEL codes: C13, C14, C15, C33.

1. INTRODUCTION

Suppose we have a fully specified model indexed by a parameter $\theta \in \Theta \subset \mathbb{R}^k$. Given a sample $Y_n = (y_1, ..., y_n)$ generated at the unknown true parameter value θ_0, a natural estimator is maximum likelihood estimator (MLE), $\hat{\theta}_{MLE} = \arg\sup_{\theta\in\Theta} \log f(Y_n|\theta)$. However, the MLE is in some situations difficult to compute due to the complexity of the model. In particular, its computation may require numerical approximations that can deteriorate the performance of the resulting approximate MLE. For example, if the model involves latent variables, they must be integrated out in order to obtain the likelihood in terms of observables. In such situations, researchers often resort to Indirect Inference (II) type methods where estimation is based on a statistic $Z_n = Z_n(Y_n)$ chosen by the researcher. This could be a set of sample moments or some other more complicated sample

Date: June 2013.

We wish to thank Manuel Arellano, Stephane Bonhomme, Federico Crudu, Ulrich Müller, Johannes Pfeifer, Peter C.B. Phillips, Enrique Sentana, Myung Seo and participants at seminars at Columbia University, Groningen University, Singapore Management University, Harvard-MIT, London School of Economics, Universidad Carlos III de Madrid, "Frontiers of Macroeconometrics" at University College London, and at the Greater New York Area Econometrics Colloquium 2010 at NYU for helpful comments and suggestions. Creel acknowledges financial support from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075), and from grant SGR2009-578. Kristensen furthermore acknowledges support from the European Research Council through a starting grant (No 312474), Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001, and the Danish Research Council through CREATES..

statistic. The standard II estimator takes the form of a continuous-updating (CU) GMM estimator,

$$\hat{\theta}_{CU} = \arg\min_{\theta\in\Theta} \frac{1}{2} \left(Z_n - E_\theta\left[Z_n\right]\right)' \Omega_n^{-1}\left(\theta\right) \left(Z_n - E_\theta\left[Z_n\right]\right), \tag{1}$$

where $\Omega_n\left(\theta\right) = E_\theta\left[\left(Z_n - E_\theta\left[Z_n\right]\right)\left(Z_n - E_\theta\left[Z_n\right]\right)'\right]$ and $E_\theta\left[\cdot\right]$ denotes expectations implied by the model evaluated at θ. In most cases, analytical expressions of $E_\theta\left[Z_n\right]$ and $\Omega_n\left(\theta\right)$ are not available, and instead simulated versions are used.[1] If Z_n is a set of sample moments, one obtains the simulated method of moments (McFadden, 1989; Duffie and Singleton, 1993). The II estimator (Gouriéroux, Monfort, Renault, 1993; Smith, 1993) chooses Z_n as an extremum estimator arriving from on an auxiliary model; in this setting, the limit of the function $\theta \longmapsto E_\theta\left[Z_n\right]$ is normally referred to as the "binding function". The efficient method of moments (Gallant and Tauchen, 1996) sets Z_n to be the score vector of an auxiliary model.

We here propose a novel II estimation method that is attractive both from a computational and statistical perspective relative to $\hat{\theta}_{CU}$. We also take as starting point some statistic Z_n , but rather than using a weighted L_2-distance to match the model with data, we propose to use the Kullback-Leibler distance. This leads to the following maximum-indirect likelihood (MIL) estimator,

$$\hat{\theta}_{MIL} = \arg\max_{\theta\in\Theta} \log f_n(Z_n|\theta). \tag{2}$$

In some situations, the above optimization problem may be difficult to solve numerically. This is particularly the case when θ is high-dimensional. To circumvent this numerical difficulty, we introduce Bayesian indirect likelihood (BIL) estimators as a computationally attractive alternative. We here focus on the posterior mean of θ given Z_n defined as

$$\hat{\theta}_{BIL} = \int_\Theta \theta f_n\left(\theta|Z_n\right) d\theta, \tag{3}$$

where, for some prior density $\pi(\theta)$ on the parameter space Θ, $f_n\left(\theta|Z_n\right)$ is the posterior distribution given by

$$f_n\left(\theta|Z_n\right) = \frac{f_n\left(Z_n,\theta\right)}{f_n\left(Z_n\right)} = \frac{f_n(Z_n|\theta)\pi\left(\theta\right)}{\int_\Theta f_n(Z_n|\theta)\pi\left(\theta\right) d\theta}.$$

We derive the asymptotic distributions of the two IL estimators and find that they are first-order equivalent to the CU version based on the same auxiliary statistic. However, due to the fact that the IL estimators utilize higher-order distributional features of Z_n that the CU version ignores, the former are shown to be higher-order efficient relative to the moment-based estimator. More precisely, higher-order expansions of the estimators reveal that the second-order variance term of the bias-adjusted IL estimators is smaller than that of the CU-II estimator. This property is a well-known feature of MLE's and generalized Bayes estimators based on the full likelihood of cross-sectional data; see, e.g., Pfanzagl and Wefelmeyer (1978) and Takeuchi and Akahira (1979). We demonstrate

[1] Alternatively, a sample-based estimator of $\Omega_n\left(\theta\right)$ can be employed but this leads to additional biases, c.f. Newey and Smith (2004).

that this fundamental result extends to our non-standard setting where the likelihood is defined in terms of a statistic rather than the full sample.

The implementation of the IL estimators requires computation of $f_n(Z_n|\theta)$. In most cases, similar to the CU-II estimator, no analytical expression of $f_n(Z_n|\theta)$ is available. We develop feasible versions of the MIL and BIL estimators by combining simulations with nonparametric density and regression techniques, respectively, as in, for example, Creel and Kristensen (2012), Fermanian and Salanié (2004), and Kristensen and Shin (2012). The simulated versions are shown to approximate the infeasible exact MIL and BIL estimators at any given tolerance level by letting the number of simulations increase sufficiently fast as $n \to \infty$. In this scenario, the simulated versions will inherit the higher-order efficiency that the exact estimators enjoy.

The above mentioned theoretical arguments for improved finite-sample performance of our IL estimators over the CU version are supported by Monte Carlo results. We investigate the performance of the proposed estimators through two examples presented in this paper: a structural auction model and a dynamic stochastic general equilibrium (DSGE) model, and additional examples presented in a longer working paper version (Creel and Kristensen, 2011). In terms of root mean squared error and bias, we find that the simulated version of the IL estimators exhibit performance that is almost always as good, and in most cases better, than the corresponding CU estimators.

There exists a related literature on Approximate Bayesian Computation (ABC); see Marin et al., 2012, for a recent survey. One implementation of ABC proposed in Beaumont, Zhang and Balding (2002) is very similar to what we call the simulated BIL (SBIL) estimator. However, this literature has mainly focused on methodology and applications in the biological sciences, including genetics, epidemiology and population biology. We here show that IL estimators are also useful in the estimation of economic models. Moreover, while the ABC literature is quite mature from an empirical point of view, only limited asymptotic theory is available for ABC estimators and their simulated versions. This paper therefore offers a number of new contributions to and extensions of what is in the ABC literature.

The remains of the paper is organized as follows: Section 2 presents the first- and higher-order theory of the estimators while Section 3 discusses the computational aspects of the estimators. Section 4 contains the simulation studies, while Section 5 concludes. All proofs and lemmas have been relegated to Appendix A and B, respectively. Appendix C provides details of the implementation of the estimators in the simulation study, while table and figures are found in Appendix D.

2. Asymptotic Properties

Let P_θ denote the family of probability measures induced by the data-generating model evaluated at $\theta \in \Theta$, and $\theta_0 \in \Theta$ the true, data-generating parameter value. We will write $P = P_{\theta_0}$. Also, let $\Phi(t)$ and $\phi(t)$ denote the cumulative distribution function (cdf) and density of a $N(0, I_d)$ distribution.

To conduct the asymptotic analysis of the IL estimators, we need to establish limit results for the indirect likelihood. This is done by making assumptions about the asymptotic behaviour of the chosen statistic, $Z_n \in \mathbb{R}^d$. Specifically, we restrict our attention to statistics that are asymptotically normally distributed around a limit $Z(\theta) \in \mathbb{R}^d$, $\sqrt{n}(Z_n - Z(\theta)) \to^d N(0, \Omega(\theta))$ under P_θ where $\Omega(\theta) \in \mathbb{R}^{d\times d}$ is the asymptotic covariance matrix. This assumption covers most known statistics in regular, stationary (in particular, cross-sectional) models by appealing to an appropriate version of the Central Limit Theorem (CLT). For convenience, we introduce the normalized statistic,

$$T_n = T(Z_n|\theta) := \sqrt{n}\Omega^{-1/2}(\theta)(Z_n - Z(\theta)),$$

which then satisfies $P_\theta(T_n \leq t) \to \Phi(t)$ as $n \to \infty$. In terms of the cdf of Z_n, $F_n(z|\theta) := P_\theta(Z_n \leq z)$, this implies that

$$F_n(z|\theta) = P_\theta(T_n \leq T_n(z|\theta)) = \Phi(T_n(z|\theta)) + o(1). \tag{4}$$

This in turn implies that $f_n(Z_n|\theta) = \sqrt{n/|\Omega(\theta)|} f_{T_n}(T_n(Z_n|\theta)|\theta)$ is well-approximated by the following sequence of Gaussian densities,

$$\phi_n^*(z|\theta) := \sqrt{n/|\Omega(\theta)|}\phi(T_n(z|\theta)). \tag{5}$$

Since $-\log \phi_n^*(Z_n|\theta)/n$ is first-order (i.e., up to order $1/\sqrt{n}$) equivalent to the quadratic loss function in eq. (1), we expect the IL estimators to be asymptotically first-order equivalent to $\hat{\theta}_{CU}$.

For the higher-order analysis, a better large-sample approximation of $f_n(Z_n|\theta)$ is needed. To this end, we assume that T_n satisfies an Edgeworth expansion of order $r \geq 0$ under P_θ,

$$f_{T_n}(t|\theta) = f_{T_n}^*(t|\theta) + R_n(t|\theta), \tag{6}$$

where the remainder term, $R_n(t|\theta)$, vanishes sufficiently fast and

$$f_{T_n}^*(t|\theta) = \phi(t)\left[1 + \sum_{i=1}^{r} n^{-i/2} a_i(t|\theta)\right], \tag{7}$$

with $t \to a_i(t|\theta)$ being a polynomial of order $3i$, $i = 1, ..., r$. The polynomial terms capture finite-sample deviations from the normal approximation; their specific forms depend on the model and the choice of Z_n. That T_n satisfies an Edgeworth expansion holds under great generality. Suppose first that the sample is i.i.d. and Z_n is a sample average, $Z_n = \sum_{i=1}^n g(y_i)/n$. Then eq. (6) holds under weak regularity conditions if $g(y_i)$ has a continuous distribution; see Hall (1992, Section 2.8). This can be extended to the case where $g(y_i)$ is discretely distributed; see, for example, Bhattacharya and Rao, 1976). If Z_n is a (sufficiently regular) estimator, the delta method can be applied in combination with the above Edgeworth expansion of sample averages to obtain that the normalized estimator, T_n, satisfies eq. (6); see, for example, Bhattacharya and Ghosh (1978), Fuh (2006), Hall and Horowitz (1996). Finally, amongst others, Phillips (1977) and Skovgaard (1981) give general conditions under which transformations of Edgeworth expandable statistics themselves have Edgeworth expansions. Thus, eq. (6) holds for a wide range of relevant statistics.

Eq. (6) implies that $f_n(z|\theta)$ should be well-approximated by

$$f_n^*(z|\theta) = \sqrt{\frac{n}{|\Omega(\theta)|}} f_{T_n}^*(T_n(z|\theta)|\theta) = \phi_n^*(z|\theta)\left[1 + \sum_{i=1}^{r} n^{-i/2} a_i(T_n(z|\theta)|\theta)\right]. \tag{8}$$

We wish to replace $f_n(z|\theta)$ by $f_n^*(z|\theta)$ in the asymptotic analysis of the IL estimators. For this to be allowed, we need to control the error

$$LR_n(\theta) := \frac{1}{n}\log\left(f_n(Z_n|\theta)/f_n^*(Z_n|\theta)\right) = \frac{1}{n}\log\left(1 + R_n(T(Z_n|\theta)|\theta)/f_{T_n}(T(Z_n|\theta)|\theta)\right).$$

In particular, the tail behaviour of $R_n(t|\theta)/f_{T_n}(t|\theta)$ has to be well-behaved. Most results on Edgeworth expansion normally provide results of the type $\sup_t R_n(t|\theta) = o\left(n^{-r/2}\right)$, but this does not suffice for our purposes and we will instead assume:

Assumption 1. *For some $r \geq 0$, $\sup_{\theta\in\Theta} E_\theta\left[\|Z_n\|^{r+2}\right] < \infty$ and eq. (6) holds with $LR_n(\theta)$ satisfying for any $c > 0$:*

$$\sup_{\theta\in\Theta} \frac{1}{n}|LR_n(\theta)| = o_P(1), \tag{9}$$

$$\sqrt{n}\frac{\partial LR_n(\theta_0)}{\partial\theta} = o_P(1), \qquad \sup_{\theta\in\Theta}\left\|\frac{\partial^2 LR_n(\theta_0)}{\partial\theta\partial\theta'}\right\| = o_P(1), \tag{10}$$

$$\sup_{\sqrt{n}\|\theta-\theta_0\|\leq c} \frac{1}{n}|LR_n(\theta)| = o_P\left(1/\sqrt{n}\right). \tag{11}$$

Eqs. (9) and (10) are used in the first-order analysis of the estimators while eq. (11) is employed in the higher-order analysis. Assumption 1 is quite high-level. However, we expect that it is satisfied under great generality. In particular, Lemma 1 in Appendix B provides a set of more primitive conditions for Assumption 1 to hold. These conditions are satisfied, for example, when Z_n is a sample average. Finally, we impose the following regularity conditions on the prior and the limiting first and second moment of Z_n:

Assumption 2. *Assume that: (i) the parameter space $\Theta \subset \mathbb{R}^k$ is compact of which the true, data-generating value θ_0 is an interior point; (ii) $\pi(\theta)$ is a continuous density with support Θ.*

Assumption 2 is completely standard. It should be noted that (ii) is only needed to develop theory for the BIL estimator.

Assumption 3. *The limit functions $\theta \mapsto Z(\theta)$ and $\theta \mapsto \Omega(\theta)$ are continuously differentiable and satisfy: (i) $Z(\theta) = Z(\theta_0)$ if and only if $\theta = \theta_0$, and (ii) $\mathcal{I}(\theta_0) := \dot{Z}(\theta_0)'\Omega^{-1}(\theta_0)\dot{Z}(\theta_0)$ has full rank where $\dot{Z}(\theta) = \partial Z(\theta)/(\partial\theta') \in \mathbb{R}^{d\times k}$.*

Assumption 3 ensures that the parameter θ_0 is identified through the statistic in the population and is similar to identification conditions for GMM-type estimators. The first part (i) ensures consistency, while the second part (ii) is used to show asymptotic normality. In particular, $\mathcal{I}^{-1}(\theta_0)$ is the asymptotic variance of the IL estimators with $\dot{Z}(\theta_0)$ capturing the information content of Z_n and $\Omega(\theta_0)$ the finite sample variation of it.

The first-order asymptotic analysis of the IL estimator now proceeds as follows: First, Assumption 1 allows us to replace the exact indirect likelihood by its first-order Edgeworth expansion. Second, for MIL, we employ standard arguments for extremum estimators in conjunction with Assumptions 2-3, to show that the maximizer of $\phi_n^*(Z_n|\theta)$, and thereby $f_n(Z_n|\theta)$, is consistent and asymptotically normally distributed. For the BIL estimator, we verify that the general results of Chernozhukov and Hong (2003) are satisfied.

Proposition 1. *Suppose that Assumptions 1-3 hold with $r \geq 1$. Then, the CU, MIL and BIL estimators are consistent and first-order equivalent, $\sqrt{n}(\hat{\theta} - \theta_0) \to^d N\left(0, \mathcal{I}^{-1}(\theta_0)\right)$, where $\hat{\theta}$ is either the CU, MIL or BIL version of the II estimator, and $\mathcal{I}(\theta_0)$ is defined in Assumption 3.*

The above result allows one to draw inference regarding the parameter. For example, confidence intervals can be computed in the standard way given an estimator of the asymptotic variance, $\mathcal{I}^{-1}(\theta_0)$. One estimator would be to utilize the sandwich form of $\mathcal{I}(\theta_0)$ as given in Assumption 3 and obtain estimates of $\dot{Z}(\theta_0)$ and $\Omega(\theta_0)$. Since these are not readily available in general, one could alternatively use $\hat{\mathcal{I}} = \frac{1}{n} \frac{\partial^2 \log f_n(Z_n|\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\hat{\theta}}$, where $\hat{\theta}$ is a consistent estimator of θ_0 such as either the MIL or BIL. This can be obtained by computing the second order derivatives of the simulated log-indirect likelihood proposed in the next section; these are available on closed form. Finally, consistent confidence bands can also be computed using the posterior quantiles. An application of Theorem 3 of Chernozhukov and Hong (2003) shows that these are valid.

For the higher-order analysis, we derive Edgeworth expansions of the three estimators and use these to show that the mean-square error (MSE) of any of the three estimators take the form $MSE\left(\sqrt{n}(\hat{\theta} - \theta_0)\right) \simeq \mathcal{I}^{-1}(\theta_0) + \Xi(\theta_0)/n + B(\theta_0) B(\theta_0)'/n^2$, where $\mathcal{I}^{-1}(\theta_0)$ is the leading variance term (common to all three estimators), while $B(\theta_0)$ is the leading bias term and $\Xi(\theta_0)$ the second-order variance component; these are specific to the particular estimator. After suitable bias adjustment, the MIL and BIL estimators will dominate the CU estimator in terms of higher-order variance in the sense that $\Xi_{CU}(\theta_0) \geq \Xi_{MIL}(\theta_0) = \Xi_{BIL}(\theta_0)$. In general, the inequality will be strict since, as demonstrated in the proof of Proposition 2 below, the higher-order terms of the MIL and BIL, as reflected by the polynomial terms of their Edgeworth expansions, differ from the ones of the CU estimator. While this is an asymptotic result, we expect that this finding translates into MIL and BIL being superior to the CU in terms of variance in finite samples.

The proof of higher-order efficiency proceeds in three steps: We first develop a higher-order stochastic expansion of each of the three estimators in terms of Z_n similar to the one in Newey and Smith (2004, Sec. 3). This in turn is used to show the distributions of the estimators satisfy Edgeworth expansions. The arguments used for the Edgeworth expansion differ from the ones usually found in the existing literature which has mainly focused on expansions under random sampling; see, for example, Bhattacharya and Ghosh (1978). In our case, we can in general not write $\log f_n(Z_n|\theta)$ as a sample average of i.i.d. variables and so the standard proof does not directly carry over to our setting. However, for the all three estimators, the higher-order expansion is a smooth function of T_n.

Since T_n satisfies an Edgeworth expansion by assumption, we can then apply the general results of Skovgaard (1981) on transformations of random sequences having an Edgeworth expansion. It now follows by standard results for likelihood-based estimators (see e.g. Pfanzagl and Wefelmeyer, 1978; Takeuchi and Akahira, 1979) that the bias-adjusted MIL and BIL estimators are second-order efficient amongst all estimators relying on the statistic Z_n, in particular, the CU estimator.

Proposition 2. *Suppose that Assumptions 2-1 hold for $r \geq 4$ with $\sup_n E\left[|Z_n|^4\right] < \infty$, and*

$$P\left(|Z_n - Z(\theta_0)| > c_1\sqrt{\log(n)/n}\right) = o\left(n^{-3/2}\right).$$

Then, the distributions of the CU-II, MIL and BIL estimators satisfy second-order Edgeworth expansions over all Borel sets A:

$$\sup_A \left| P\left(\sqrt{n}\mathcal{I}(\theta_0)\left(\hat{\theta} - \theta_0\right) \in A\right) - \int_A \phi(x)\left[1 + \sum_{i=1}^{2} n^{-i/2}\tilde{a}_i(x)\right]dx \right| = o\left(n^{-1}\right),$$

where $\tilde{a}_i(x)$ is a polynomial of order $3i$, $i = 1, 2$, whose coefficients depend on the particular estimator. In particular, the bias-adjusted versions of the MIL and BIL are second-order equivalent and efficient relative to any other estimator based on Z_n, including the CU-II.

This shows that MIL and BIL estimators are second-order equivalent when adjusted for their leading bias terms. That is, the polynomial terms corresponding to the variance terms are identical for the two estimators. If the first-order derivative of the polynomial $a_1(t|\theta)$ appearing in the Edgeworth expansion in eq. (6) is equal to zero then we find that the CU estimator will be second-order equivalent to the IL estimators. However, in general, the CU estimator will in general have a different variance component and so not enjoy second-order efficiency.

Expressions of the leading bias term, $B(\theta_0)$, for each of the three estimators are provided in Lemma 2 in Appendix B. From these expressions, we see that the three estimators share a common bias component, $B_I(\theta)$, which corresponds to the one of the empirical likelihood estimator reported in Newey and Smith (2004, Theorem 4.6). Each of the three estimators has an additional bias component: The MIL estimator has $B_{MIL}(\theta)$ which is caused by the higher-order curvature of the indirect log-likelihood as captured by the second-order derivative of $a_1(t|\theta)$. The CU estimator contains an additional bias component due to curvature of the covariance matrix while the bias of the additional bias of the BIL estimator are partially caused by the prior. Note that the bias component induced by the prior vanishes if the prior is chosen as a uniform density. In the just-identified case ($\dim(Z_n) = \dim(\theta)$), $\Omega^{-1}(\theta)$ can be chosen as the identity matrix and some of the bias terms drop out. Gouriéroux, Renault and Touzi (2000) and Gouriéroux, Phillips and Yu (2010) have advocated using II for bias adjustment: Given an initial, potentially biased, estimator $\tilde{\theta}$, we choose $Z_n = \tilde{\theta}$ with the resulting II estimator having reduced bias. In this case, $\ddot{Z}(\theta) = 0$ and so $\hat{\theta}_{CU}$ has no first-order bias while the MIL and BIL estimators maintain some biases due to higher-order curvature of the log-likeliood and the prior. However, through extensive simulations reported in the working paper version (Creel and Kristensen, 2011), we found that when II is used for bias adjustment,

MIL and BIL performed just as well, and most of the time better, compared to the traditional CU version.

3. Simulated Versions of MIL and BIL

In most situations, the likelihood $f_n(Z_n|\theta)$ will not be available, and one has to resort to numerical approximations instead. We here propose easy-to-compute simulated versions of the MIL and BIL. Since the model is simulable and the mapping $Z_n(\theta) \equiv Z_n(Y_n(\theta))$ is known, we can draw S independent samples, $Y_n^s(\theta)$ for $s = 1, ..., S$, from the model evaluated at the trial value θ, and compute the associated statistic, $Z_n^s(\theta) \equiv Z_n(Y_n^s(\theta))$, $s = 1, ..., S$. These simulated versions of the statistic are then fed into a kernel density estimator (see e.g. Li and Racine, 2007, Ch. 1 for an introduction):

$$\hat{f}_{n,S}(Z_n|\theta) = \frac{1}{S}\sum_{s=1}^{S} K_h\left(Z_n^s(\theta) - Z_n\right), \tag{12}$$

where $K_h(z) = K(z/h)/h$, $K(z)$ is a kernel function and $h > 0$ is a bandwidth. We embed the simulated density inside (2) yielding a simulated MIL (SMIL) estimator. For proof technical reasons, we have to trim the tails of the simulated likelihood and so the formal definition of the SMIL takes the form

$$\hat{\theta}_{SMIL} = \arg\max_{\theta\in\Theta} \hat{\tau}_a(Z_n|\theta) \log \hat{f}_{n,S}(Z_n|\theta), \tag{13}$$

where $\hat{\tau}_a(z|\theta)$ is a trimming function satisfying $\hat{\tau}_a(z|\theta) = 1$ for $\hat{f}_{n,S}(z|\theta) > a$ and $\hat{\tau}_a(z|\theta) = 0$ for $\hat{f}_{n,S}(z|\theta) < a/2$, for some trimming parameter $a > 0$. The introduction of trimming is only used to establish certain theoretical properties of the simulated MIL estimator; in practice, for reasonable large number of simulations, trimming can be left out; this is the case for our simulation study. The simulated MIL estimator is akin to the nonparametric simulated maximum-likelihood estimator (NPSMLE) of Fermanian and Salanié (2004) and Kristensen and Shin (2012).

For the computation of the BIL estimator, we also combine simulations and nonparametric techniques: Make i.i.d. draws θ^s, $s = 1, ..., S$, from the pseudo-prior density $\pi(\theta)$, for each draw generate a sample $Y_n(\theta^s)$ from the model at this parameter value, and then compute the corresponding statistic $Z_n^s = Z(Y_n(\theta^s))$, $s = 1, ..., S$. Given the i.i.d. draws (θ^s, Z_n^s), $s = 1, ...S$, we can obtain a simulated version of the BIL (SBIL) through nonparametric regression techniques. One such is the kernel estimator (see Li and Racine, 2007, Ch. 2),

$$\hat{\theta}_{SBIL} = \hat{E}_S[\theta|Z_n] = \frac{\sum_{s=1}^{S} \theta^s K_h\left(Z_n^s - Z_n\right)}{\sum_{s=1}^{S} K_h\left(Z_n^s - Z_n\right)}. \tag{14}$$

Alternatively, the k-nearest neighbor (KNN) estimator (see Li and Racine, 2007, Ch. 14) can be used. This takes the form above but now the bandwidth is chosen as $h = d_k(Z_n)$ with $d_k(Z_n)$ denoting the Euclidean distance between Z_n and the k-th nearest neighbor among the simulated values. The idea of combining simulations with kernel regressions also appears in Creel and Kristensen (2012).

For both the SMIL and SBIL, there are two additional sources of error relative to the exact MIL and BIL estimators: Randomness is added due to the use of simulations, and there is additional biases due to kernel smoothing. However, as $h \to 0$ and $Sh^d \to \infty$, the nonparametric density and regression estimators converge towards the exact likelihood and posterior mean, respectively, for any given sample size. Thus, using standard bandwidth selection rules, both errors can be controlled for by choosing S sufficiently large. Note that the use of kernel regression and simulations in the computation of BIL is computationally feasible because $\dim(Z_n)$ is fixed and relatively small. In contrast, it would be computationally infeasible to use similar techniques to compute the full Bayesian posterior mean $E(\theta|Y_n)$ since $\dim(Y_n)$ in most applications is prohibitively large.

In the ABC literature, as discussed in the introduction, similar methods for computing the BIL estimator have been suggested. Some of these employ more advanced samplers such as Markov chain Monte Carlo, sequential Monte Carlo and importance sampling (the survey by Marin et al., 2012, provides references). The resulting simulated estimators should in principle be computationally more efficient compared to the one in equation (14). But these on the other hand require more careful implementation; otherwise, they might perform poorly and even fail to converge (see, e.g., Kormiltsina and Nekipelov, 2009). To avoid any such issues, we focus on the basic sampler proposed above in the simulation study.

For the analysis of the simulated versions, we impose the following additional assumptions on the model and the chosen statistic:

Assumption 4. *For all $n \geq 1$ and $\theta \in \Theta$, $z \mapsto f_n(z|\theta)$ is a density with respect to the Lebesgue measure, twice continuously continuously differentiable in z, and is bounded away from zero on any compact set. Moreover, $Z_n(\theta)$ is continuously differentiable w.r.t. θ and satisfies $\sup_{n\geq 2} E\left[\sup_{\theta\in\Theta} \|Z_n(\theta)\|^q\right] < \infty$ for some $q > 0$.*

This assumption should be satisfied for most regular models that are smooth in θ, and for standard choices of statistics. We restrict the kernel to satisfy:

Assumption 5. *The kernel $K(z)$ is differentiable with $sup_z |K'(z)| < \infty$ and, for some $a > 1$, $|K(z)| \leq C\|z\|^{-a}$ for $\|z\| > L$. Moreover, $\int K(z)\,dz = 1$, $\int zK(z)\,dz = 0$, $\int z^2K(z)\,dz = 1$.*

Under these assumptions on the kernel and the model, the following result holds:

Proposition 3. *Assume that Assumptions 2-5 hold. Then the SMIL and SMIL are equivalent to the MIL and BIL, respectively, up to order $n^{-r/2}$, $r \geq 1$, if, with $d = dim(Z_n)$, and $q > 0$ given in Assumption 4,*

$$SMIL \quad : \quad a^{-1}n^{r-1/2}h^2 \to 0 \; a^{-2}n^{2r-d-1}\log S/\left(Sh^d\right) \to 0 \text{ and } n^{r-1}(n/a)^{-(q-1)/(2nq)} \to 0.$$

$$SBIL \quad : \quad n^{r+1}h^2 \to 0 \text{ and } n^{2r+d}/\left(Sh^d\right) \to 0.$$

The requirements on the bandwidth and number of simulations are used to control additional biases and variances due to simulations and kernel smoothing as sample size grows. They are somewhat different compared to the ones found in, for example, Kristensen and Shin (2012). The reason for this is that the target density, $f_n(z|\theta)$, asymptotically has a pole at $z = Z(\theta)$ which leads to non-standard behaviour of $\hat{f}_{n,S}(z|\theta)$ in large

samples. However, for fixed sample size, the bias and variances of $\hat{f}_{n,S}(z|\theta)$ are completely standard as known from the literature on kernel estimation. In particular, both SMIL and SBIL suffer from the usual curse of dimensionality associated with nonparametric methods. This appears explicitly in the conditions on S and h given in Proposition 3 where as d (which must be at least that of θ, and which is larger in most of the applications below) increases, we have to use a large number of simulations in order to control the stochastic approximation error.

The bandwidth should be chosen differently depending on whether the SMIL or SBIL estimators are employed. For SBIL, one can use standard bandwidth selection methods for kernel regression estimators such as cross-validation or plug-in methods. For SMIL, log-likelihood based likelihood-based cross-validation (Hall, 1987) could be used or, in large samples where $f_n(z|\theta)$ is well-approximated by a normal density, Silverman's Rule of Thumb.

In practice, we choose the number of simulations S so large, that the additional variance due to simulations is negligible. However, for completeness, we note that the simulated version of the BIL estimator satisfies $\hat{\theta}_{SBIL} = \hat{\theta}_{BIL} + E_S[Z_n]$, for a stochastic function $E_S(z)$ which only depends on the simulations. In the case of kernel regression, $\sqrt{Sh^d}E_S(Z_n) \to^d N\left(0, \|K\|^2 \sigma_n^2(Z_n)/f_n(Z_n)\right)$, as $Sh^d \to \infty$, where $\|K\|^2 = \int K^2(z)\,dz$, and $\sigma_n^2(Z_n) = Var[\theta|Z_n]$. Thus, the variance estimator of the kernel-smoothed version of SBIL could be adjusted by adding $\|K\|^2 \frac{\sigma_n^2(Z_n)}{f_n(Z_n)}/\left(Sh^d\right)$ to $J^{-1}(\theta_0)$. Similar results hold for the nearest-neighbor version. Using the arguments of Kristensen and Salanié (2010), standard errors of SMIL can be adjusted in a similar fashion.

4. Monte Carlo Results

In this section we explore the finite sample performance of the SMIL and SBIL estimators in the context of a structural auction model and a DSGE model. Creel and Kristensen (2011) contains additional examples, including simple time series models, dynamic and nonlinear panel data models. Due to space constraints, precise details of the implementation, including choice of prior and auxiliary statistics, are provided in Appendix C. All software used to compute the results is free, and complete code and all software required to replicate all results reported in this paper is available from the authors.

4.1. Auction Model.

Li (2010) proposes to use II for estimation of structural econometric models, and illustrates with a Monte Carlo example of estimation of the parameters of a Dutch auction, where only the winning bid is observed. We observe n i.i.d. auctions. At each auction $i = 1, 2, ..., n$, the quality, x_i, of the item being auctioned is observed; this follows a uniform $(0, 1)$ distribution. Given this signal, N agents make a bid based on their private value of the item. Their privates values are mutually independent and come from a common exponential distribution with mean $\exp(\theta_0 + \theta_1 x_i)$. The equilibrium strategy for the winning bid is then $b_i^* = v_i^* - \int_0^{v_i^*} F^{N-1}(u|x_i)du / F^{N-1}(v_i^*|x_i)$ where v_i^* is the highest private valuation, and $F(\cdot|x_i)$ is the exponential distribution function. For a given value of N, symbolic computation software can be used to obtain an analytic solution for the winning bid, so simulations can be generated very quickly. The observed

data are the n values of $\{x_i, b_i^*\}$, and we seek to estimate θ_0 and θ_1. We set $N = 6$ and the true parameter values to $\theta_0 = 0.5$ and $\theta_1 = 0.5$. For BIL, we choose the prior as uniform over $(-1, 3) \times (0, 2)$. We introduce an auxiliary model $\log b_i^* = \beta_0 + \beta_1 x_i + \epsilon_i$, and choose Z_n as the OLS estimates of (β_0, β_1) together with the log-residual variance and the first three central moments of the logarithm of the winning bid. Thus, we have six statistics to identify the two parameters. Given the experimental design, the chosen statistics are notably non-normally distributed, so we can expect to see differences between the IL and CU-II estimators.

Table 1 contains the results for 5000 Monte Carlo replications, comparing SMIL, SBIL and CU-II. For samples of size 80, for which we have results for all three estimators, we see that the SBIL estimator obtains lower RMSEs for both parameters than does SMIL. Based on this, and on additional results in Creel and Kristensen (2011) where the SBIL estimator performs as well or better than SMIL, we focus henceforth on the SBIL estimator, which is computationally considerably more convenient. Comparing the SBIL estimator to CU-II, we see that SBIL has a low bias for all sample sizes, while the CU-II estimator has a more notable bias for the smallest sample size. Comparing RMSEs, SBIL has a considerable advantage, though the gap narrows somewhat as the sample size increases, in line with the first order asymptotic equivalence of the two estimators. This example shows that there can be considerable gains over CU-II, due to the ability of the SBIL estimator to take into account the small sample features of the statistic's distribution.

4.2. **DSGE Model.** Next, we report results for estimation of a simple DSGE model. Full likelihood-based estimation is complicated by two issues: First, DSGE models often contain unobserved state variables that have to be filtered out in the computation of the likelihood. In a linearized version, the Kalman filter can be employed while for higher-order solutions, nonlinear filtering methods can be computationally challenging. Second, standard models as a rule contain fewer shocks than state variables, which leads to stochastic singularities. This can be resolved in two ways: First, by adding on measurement errors to the observable variables, but this solution involves integrating out additional variables and the artificially introduced noise leads to less precise inference. Second, by using at most as many observable variables as structural shocks, but this will in general impart an efficiency loss. These two challenges complicate full likelihood-based estimation and inference considerably; see, amongst others, Fernández-Villaverde and Rubio-Ramírez (2005), An and Schorfheide (2006), Fernández-Villaverde (2010) and Winschel and Krätzig (2010) for recent discussions.

As we shall see, IL avoids the above issues. First of all, there is no need for filtering; we only need to simulate the chosen statistic. Second, stochastic singularities can be avoided by choosing individual statistics that are not perfectly multicollinear. In particular, it is possible to find independent moments that incorporate information about more variables than those that are linearly independent. This makes it an attractive alternative both from a computational and statistical perspective. Our approach is related to Ruge-Murcia (2012) who employs simulated method of moments (SMM) for the estimation of a

DSGE model. Recall that SMM is a special case of the CU-II estimator, and so, in comparison with IL, his method requires numerical optimization, which can be computationally demanding when the parameter space is large, and is expected to be less efficient in finite samples. In a simulation study, Ruge-Murcia (2012) treats a number of the parameters as known; in contrast, we here estimate all parameters entering the model.

The model that we consider is as follows: A single good can be consumed or used for investment, and a single competitive firm maximizes profits. The states variables are: y output; c consumption; k capital; i investment, n labor; w real wages; r return to capital. The household maximizes expected discounted utility

$$E_t \sum_{s=0}^{\infty} \beta^s \left(\frac{c_{t+s}^{1-\gamma}}{1-\gamma} + (1 - n_{t+s}) \eta_t \psi \right)$$

subject to the budget constraint $c_t + i_t = r_t k_t + w_t n_t$ and the accumulation of capital $k_{t+1} = i_t + (1 - \delta k_t)$. There is a preference shock, η_t, that affects the desirability of leisure. The shock evolves according to $\ln \eta_t = \rho_\eta \ln \eta_{t-1} + \sigma_\eta \epsilon_t$. The competitive firm produces the good y_t using the technology $y_t = k_t^\alpha n_t^{1-\alpha} z_t$. Technology shocks z_t also follow an AR(1) process in logarithms: $\ln z_t = \rho_z \ln z_{t-1} + \sigma_z u_t$. The innovations to the preference and technology shocks, ϵ_t and u_t, are mutually independent i.i.d. standard normally distributed. The good y_t can be allocated by the consumer to consumption or investment: $y_t = c_t + i_t$. The consumer provides capital and labor to the firm, and is paid at the rates r_t and w_t, respectively. The unknown parameters are collected in $\theta = (\alpha, \beta, \delta, \gamma, \psi, \rho_z, \rho_\eta, \sigma_z, \sigma_\eta)$. In total, we have seven state variables and only two shocks.

In the estimation, we treat capital stock k as unobserved while the remaing state variables are observed. Two sample sizes are used: $n = 40$ and $n = 160$, which mimic 10 and 40 years of quarterly data, respectively. We explore two different designs for the true parameters, which are given in the second columns of Tables 3 and 4. Both designs set true steady state hours to 1/3 of the time endowment. Apart from that, the first design sets the parameters to values that are intended to be typical of the DSGE literature, while the second design uses less typical values to check the ability of the SBIL estimator to detect departures from the usually encountered values. For example, the discount rate is somewhat low, and the depreciation of capital is somewhat high. The shocks are more volatile in the second design.

Traditionally, DSGE models are estimated by linearizing the model and assuming Gaussian shocks such that the Kalman filter can be employed; see, e.g., Smets and Wouters (2007). As a first step, we therefore attempted to estimate θ in a linearized version of the model using full likelihood methods. This was done by computing the Bayesian posterior using MCMC and Kalman filtering within the Dynare software package (Adjemian et al., 2011). Due to the stochastic singularities, we used only two variables, c and n, to compute the likelihood. However, estimation in this manner was not possible since no posterior mode existed. This is consistent with previous findings of weak identification in linearized models (Canova and Sala, 2009; Fernández-Villaverde and Rubio-Ramírez, 2005; Iskrev, 2010).

We therefore in the following focus on a third-order perturbation solution, that combines good accuracy with moderate computational demands (Aruoba, Fernández-Villaverde and Rúbio-Ramírez, 2006), and SBIL. For both designs, the prior for SBIL is the same. Our pseudo-prior $\pi(\theta)$ is a uniform distribution over the hypercube defined by the bounds of the parameter space, which are found in Table 2. The chosen bounds cause the pseudo-prior means to be biased for the true parameter values, as one may note in Figure 1, which is discussed further below. The bounds are intended to be broad, in comparison to the fairly strongly informative priors that are often used when estimating DSGE models (Fernández-Villaverde, 2010, discusses use of strongly informative priors).

We use two versions of the SBIL estimator. The first version is the one presented and analyzed in Sections 2-4. It contains $d = 24$ elements, composed of coefficients of auxiliary regressions, sample means, sample autocovariances, etc (details arg given in Appendix C and in the example code). While the asymptotic tells us that the BIL using the entire vector of statistics will yield the best estimates, we also know that the simulated version will suffer from a curse of dimentionality, necessitating use of a large number of simulations. To explore the possibility of obtaining accurate results without using an exceptionally large number of simulations, we also report results where the posterior mean for each parameter is computed using only a subset of the full set of sample statistics (details of this "targeting" approach are given in Appendix C). Developing an asymptotic theory of this "targeting" version of the IL estimator is left for future research.

Tables 3 and 4 give the SBIL estimation results, for the two designs, respectively. All parameters are estimated with good precision, especially for the larger sample size. We see that the targeted SBIL estimator always gives more precise results than does the simple estimator that uses the same auxiliary statistic for all parameters, except in the case of ρ_z, ρ_η, for the small sample size. Comparing across sample sizes, RMSE roughly halves when the sample size quadruples, in line with the first order asymptotic theory. When there is a non-negligible bias for the small sample size (for example, the parameters ρ_z, ρ_η and ψ), it declines notably when the sample size increases. The combination of bias reduction compared to the prior and contraction of RMSE illustrates that all parameters of the model are well-identified by the chosen auxiliary statistics. Figure 1 shows the pseudo-prior densities, true parameter values, and a kernel density estimate of the sampling density of the targeted SBIL estimator, fit using the 5000 Monte Carlo replications, for each parameter of the first design. We see that the density of the SBIL estimator moves toward and concentrates about the true parameter value, even for a pseudo-prior distribution that is biased (the true parameter values are not in the center of the plots) and quite uninformative (the pseudo-priors take low values compared to the peak of the density of the SBIL estimator).

Given the simplicity and good performance of the SBIL for estimation of the DSGE model, we believe that it provides an interesting alternative to the considerably more complex and computationally demanding methodology of MCMC combined with particle filtering, which can probably be described as the current state of the art for estimation of DSGE models. Our results are suggestive that inference though a statistic may not (depending on the statistic, of course) be subject to identification problems when the model

is solved using a more accurate higher order solution method, rather than linearized. Moreover, possible stochastic singularities do not cause any difficulties for SBIL estimation, except that one might need to choose auxiliary statistics with some care to avoid inflating the dimension of the auxiliary statistic with additional elements that add no new information. A measure of the practicality of estimation of a DSGE model using SBIL is the simple fact that we have been able to perform 20,000 Monte Carlo replications of SBIL: 5000 replications for each of two designs and for each of two sample sizes. We are aware of no similar Monte Carlo exploration of the MCMC/SMC and particle filtering combination for estimation of DSGE models.

5. Conclusions

This paper has introduced IL (also known as ABC) estimators to the econometric literature. We analyzed the asymptotic properties of the estimators and showed that they are higher-order efficient relative to the corresponding standard CU-II estimator. Simulation studies showed that indeed the IL estimators enjoy good properties in finite samples, with small biases and variances. From a methodological point of view, a number of extensions would be of interest: For example, developing tools for selection of a suitable statistic Z_n for a given model; see Fearnhead and Prangle (2012) for some results in this direction. In addition, developing and analyzing inferential tools using the indirect likelihood, such as likelihood ratio tests, would be useful. There is also scope for improvements in terms of numerical implementation. We have focused on the basic sampler as given in eq. (14) choosing the number of neighbors k through the simple rule $k = a \times S^{0.25}$. More sophisticated rules, such as cross-validation, or different kernels, could lead to better performance. Similarly, more complicated samplers using importance sampling methods could be used to improve on the computation time.

References

[1] Adjemian, S., H. Bastani, M. Juillard, F. Karamé, F. Mihoubi, G. Perendia, J. Pfeifer, M. Ratto and S. Villemot, 2011, "Dynare: Reference Manual, Version 4," Dynare Working Papers, 1, CEPREMAP.

[2] An, S. and F. Schorfheide, 2007, "Bayesian Analysis of DSGE Models", *Econometric Reviews*, 26, 113-172.

[3] Aruoba, S.B, J. Fernández-Villaverde and J. Rubio-Ramírez, 2006, "Comparing solution methods for dynamic equilibrium economies", *Journal of Economic Dynamics & Control*, 30, 2477-2508.

[4] Beaumont, M., W. Zhang and D. Balding, 2002, "Approximate Bayesian computation in population genetics", *Genetics*, 162, 2025-2035.

[5] Bhattacharya, R.N. and J.K. Ghosh, 1978, "On the validity of the formal Edgeworth Expansion," *Annals of Statistics* 6, 434-451.

[6] Bhattacharya, R.N. and R. R. Rao, 1976, *Normal Approximations and Asymptotic Expansions*. New York: Wiley.

[7] Canova, F. and L. Sala, 2009, "Back to square one: Identification issues in DSGE models," *Journal of Monetary Economics*, 56, 231-249.

[8] Chernozhukov, V. and H. Hong, 2003, "An MCMC approach to classical estimation," *Journal of Econometrics* 115, 293-346.

[9] Creel, M. and D. Kristensen, 2011, "Indirect likelihood inference," Working Paper 558, Barcelona Graduate School of Economics.

[10] Creel, M. and D. Kristensen, 2012, "Estimation of dynamic latent variable models using simulated nonparametric moments," *Econometrics Journal* 15, 490-515.

[11] Duffie, D. and K. J. Singleton, 1993, "Simulated moments estimation of Markov models of asset prices," *Econometrica*, 61, 929–952.

[12] Fearnhead, P. and Prangle, D., 2012, "Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation", *Journal of the Royal Statistical Society: Series B (Statistical Methodology*), 74, 419–474.

[13] Fermanian, J.-D. and B. Salanié, 2004, "A nonparametric simulated maximum likelihood estimation method," *Econometric Theory*, 20, 701-734.

[14] Fernández-Villaverde, J, 2010, "The econometrics of DSGE models," *SERIEs*, 1, 3–49.

[15] Fernández-Villaverde, J. and J. Rubio-Ramírez, 2005, "Estimating dynamic equilibrium economies: linear versus nonlinear likelihood," *Journal of Applied Econometrics*, 20, 891– 910.

[16] Fuh, C.-D., 2006, "Efficient likelihood estimation in state space models," *Annals of Statistics* 34, 2026-2068.

[17] Gallant, A. R. and G. Tauchen, 1996, "Which moments to match?" *Econometric Theory* 12, 657-681.

[18] Gouriéroux, C., A. Monfort, and E. Renault, 1993, "Indirect inference," *Journal of Applied Econometrics*, 8, S85-S118.

[19] Gouriéroux, C., P.C.B. Phillips and J. Yu, 2010, "Indirect inference for dynamic panel models," *Journal of Econometrics* 157, 68-77.

[20] Gouriéroux, C., E. Renault and N. Touzi, 2000, "Calibration by simulation for small sample bias correction," in Mariano, R.S., Schuermann, T., Weeks, M. (Eds.), *Simulation-Based Inference in Econometrics: Methods and Applications*, pp. 328-358. Cambridge: Cambridge University Press.

[21] Hall, P., 1987, "On Kullback-Leibler loss and density estimation," *Annals of Statistics* 25, 1491-1519.

[22] Hall, P., 1992, *The Bootstrap and Edgeworth Expansion*, New York: Springer.

[23] Hall, P. and J.L. Horowitz, 1996, "Bootstrap critical values for tests based on Generalized-Method-of-Moments estimators," *Econometrica* 64, 891-916.

[24] Iskrev, N., 2010, "Local identification in DSGE models," *Journal of Monetary Economics,* 57, 189–202.

[25] Johnson, R., 1970, "Asymptotic expansions associated with posterior distributions," Annals of Mathematical Statistics 41, 851-864.

[26] Kormiltsina, A. and D. Nekipelov, 2009, "Numerical performance of MCMC algorithms for classical estimation," working paper, UC Berkeley.

[27] Kristensen, D., 2009, "Uniform convergence rates of kernel estimators with heterogeneous, dependent data," *Econometric Theory* 25, 1433-1445.

[28] Kristensen, D. and B. Salanié, 2010, "Higher order improvements for approximate estimators," CAM Working Papers 2010-04, University of Copenhagen.

[29] Kristensen, D. and Y. Shin, 2012, "Estimation of dynamic models with nonparametric simulated maximum likelihood," *Journal of Econometrics* 167, 76–94.

[30] Li, Q. and J. Racine, 2007, *Nonparametric Econometrics: Theory and Practice*. Princeton: Princeton University Press.

[31] Li, T., 2010, "Indirect inference in structural econometric models," *Journal of Econometrics* 157, 120-128.

[32] Marin, J.-M., P. Rudlo, C. Robert and R. Ryder, 2012, "Approximate Bayesian computational methods, *Statistical Computing*, 22, 1167-1180.

[33] McFadden, D., 1989, "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica*, 57, 995–1026.

[34] Newey, W.K. and D. McFadden, 1994, "Large sample estimation and hypothesis testing," in: R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Vol. IV, 2111-2245. Amsterdam: Elsevier Science.

[35] Newey, W. K. and R. J. Smith, 2004, "Higher order properties of GMM and generalized empirical likelihood estimators," *Econometrica*, 72, 219–255.

[36] Pfanzagl, J. and W. Wefelmeyer, 1978, "A third-order optimum property of the maximum likelihood estimator," *Journal of Multivariate Analysis*, 8, 1-29.

[37] Phillips, P.C.B, 1977, "A General Theorem in the Theory of Asymptotic Expansions as Approximations to Finite Sample Distributions of Econometric Estimators," *Econometrica* 45, 1517- 1534.

[38] Ruge-Murcia, F., 2007, "Methods to estimate dynamic stochastic general equilibrium models," *Journal of Economic Dynamics and Control*, 31, 2599-2636.

[39] Ruge-Murcia, F., 2012, "Estimating nonlinear DSGE models by the simulated method of moments: With an application to business cycles," *Journal of Economic Dynamics and Control*, 36, 914–938.

[40] Skovgaard, I., 1981, "Transformation of an Edgeworth expansion by a sequence of smooth functions," *Scandinavian Journal of Statistics* 8, 207-217.

[41] Smets F. and R. Wouters, 2007, "Shocks and frictions in US business cycles: a Bayesian DSGE approach," *American Econonomic Review,* 97, 586–606.

[42] Smith, A., 1993, "Estimating nonlinear time series models using simulated vector autoregressions," *Journal of Applied Econometrics*, 8, S63-S84.

[43] Takeuchi, K. and M. Akahira (1979) "Asymptotic optimality of the generalized Bayes estimator in multiparameter case," *Annals of the Institute of Statistical Mathematics* 31, Part A, 403-415.

[44] Taniguchi, M., 1987, "Validity of Edgeworth expansions of minimum contrast estimators for Gaussian ARMA processes," *Journal of Multivariate Analysis* 21, 1–28

[45] Winschel, V. and Krätzig, M., 2010, "Solving, estimating, and selecting nonlinear dynamic models without the curse of dimensionality," *Econometrica*, 78, 803–821.

APPENDIX A: PROOFS

Proof. [Proposition 1] We first show consistency of the MIL estimator. Due to eq. (9), we can treat $\frac{1}{n}\log\phi_n^*(Z_n|\theta)$ as the actual log-likelihood. This satisfies, uniformly in $\theta \in \Theta$,

$$\begin{aligned}\frac{1}{n}\log\phi_n^*(Z_n|\theta) &= -\frac{1}{2n}\log\left(|\Omega(\theta)|\right) - \frac{1}{2n}\left(Z_n - Z(\theta)\right)'\Omega^{-1}(\theta)\left(Z_n - Z(\theta)\right) + o_P(1) \\ &= -\frac{1}{2}\left(Z(\theta_0) - Z(\theta)\right)'\Omega^{-1}(\theta)\left(Z(\theta_0) - Z(\theta)\right) + o_P(1). \qquad (15)\end{aligned}$$

The limit is a continuous function w.r.t. θ with a unique minimum at $\theta = \theta_0$ by Assumption 3. It now follows by standard results (see e.g. Newey and McFadden, 1994, Theorem 2.1), that the MILE is consistent. Next, observe that

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial\log f_n(Z_n|\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} &= \frac{1}{\sqrt{n}}\frac{\partial\log\phi_n^*(Z_n|\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} + o_P(1) \\ &= \dot{Z}(\theta_0)'\Omega^{-1}(\theta_0)\sqrt{n}\left(Z_n - Z(\theta_0)\right)' + o_P(1) \to^d N\left(0, \mathcal{I}^{-1}(\theta_0)\right)\end{aligned}$$

while, uniformly in θ,

$$\begin{aligned}\frac{1}{n}\frac{\partial^2\log f_n(Z_n|\theta)}{\partial\theta\partial\theta'} &= \frac{1}{n}\frac{\partial^2\log\phi_n^*(Z_n|\theta)}{\partial\theta\partial\theta'} + o_P(1) = \ddot{Z}(\theta)'\Omega^{-1}(\theta_0)\left(Z_n - Z(\theta)\right)' + \mathcal{I}(\theta) + o_P(1) \\ &= \ddot{Z}(\theta)'\Omega^{-1}(\theta_0)\left(Z(\theta_0) - Z(\theta)\right)' + \mathcal{I}(\theta) + o_P(1).\end{aligned}$$

Asymptotic normality of the MIL estimator now follows by standard arguments for extremum estimators (see e.g. Newey and McFadden, 1994, Theorem 2.1).

The first-order properties of the BIL are established by verifying Assumptions 1-4 in Chernozhukov and Hong (2003), CH henceforth. First note that CH's Assumptions 1-2 are satisfied by our Assumption 2. What remains is to verify their Assumption 3-4. But by combining their Lemmas 1-2 with the above derivations, these are easily verified. We can now appeal to CH's Theorem 2 which yields the desired result.

Finally, using that $E_\theta[Z_n] = Z(\theta) + o(1/n)$ and $\Omega_n(\theta) = \Omega(\theta) + o(1/n)$, the properties of the CU version follows along the same lines as for the MIL estimator. □

Proof. [Proposition 2] First consider the MIL estimator. Lemma 4 together Assumption 1 imply that we can treat the higher-order Edgeworth expansion as the actual log-likelihood. We then employ Lemma 3 with $m_n(\theta) := \frac{1}{n}\partial\log f_n^*(Z_n|\theta) / (\partial\theta)$ and

$$Dm(\theta) := -\mathcal{I}(\theta), \; D^2m(\theta) := -3\dot{Z}(\theta)'\left\{\Omega^{-1}(\theta)\ddot{Z}(\theta) + \frac{\partial\Omega^{-1}(\theta)}{\partial\theta}\dot{Z}(\theta)\right\}, \qquad (16)$$

to obtain a higher-order expansion of the MIL estimator. Write $\log f_n^*(Z_n|\theta) = \log\phi_n^*(Z_n|\theta) + \log g_n(Z_n|\theta)$ where $g_n(Z_n|\theta) := 1 + \sum_{i=1}^r n^{-i/2}a_i(T_n(\theta)|\theta)$. contains the higher-order terms. The Gaussian component of the score satisfies, with $\Delta_n(\theta) = \sqrt{n}(Z_n - Z(\theta)) = \Omega^{1/2}(\theta)T_n(\theta)$,

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial\log\phi_n^*(Z_n|\theta)}{\partial\theta} &= \dot{Z}(\theta)'\Omega^{-1/2}(\theta)T_n(\theta) + \frac{1}{2\sqrt{n}}\frac{\partial\log\left|\Omega^{-1}(\theta)\right|}{\partial\theta} \\ &\quad -\frac{1}{2\sqrt{n}}T_n(\theta)'\Omega^{1/2}(\theta)\frac{\partial\Omega^{-1}(\theta)}{\partial\theta}\Omega^{1/2}(\theta)T_n(\theta)\end{aligned}$$

while the one of the Hessian satisfies

$$\begin{aligned}\frac{1}{\sqrt{n}}\left\{\frac{\partial^2 \log \phi_n^*(Z_n|\theta)}{\partial\theta\partial\theta'} - Dm(\theta)\right\} &= \left[\ddot{Z}(\theta)'\Omega^{-1/2}(\theta) + 2\dot{Z}(\theta)'\frac{\partial\Omega^{-1}(\theta)}{\partial\theta}\Omega^{1/2}(\theta)\right]T_n(\theta) \\ &+ \frac{1}{2\sqrt{n}}T_n(\theta)'\Omega^{1/2}(\theta)\frac{\partial^2\Omega^{-1}(\theta)}{\partial\theta^2}\Omega^{1/2}(\theta)\,T_n(\theta) \\ &+ \frac{1}{2\sqrt{n}}\frac{\partial^2 \log\left|\Omega^{-1}(\theta)\right|}{\partial\theta^2}\end{aligned}$$

and similarly for the third derivative. The higher-order component of the score satisfies

$$\frac{\partial \log g_n(Z_n|\theta)}{\partial\theta} = \frac{\sum_{i=0}^{r} n^{-i/2} a_{1,i}(T_n(\theta)|\theta)}{g_n(Z_n|\theta)},$$

where $a_{1,i}(t|\theta)$ is a polynomial; in particular, $a_{1,0}(t|\theta) := -a_1^{(1)}(t|\theta)\,\Omega^{-1/2}(\theta)\,\dot{Z}(\theta)$. Similarly, again by collecting polynomials with the same order,

$$\frac{\partial^2 \log g_n(Z_n|\theta)}{\partial\theta\partial\theta'} = \sqrt{n}\frac{\sum_{i=0}^{2r-1} n^{-i/2} a_{2,i}(T_n(\theta)|\theta)}{g_n(Z_n|\theta)},$$

where $a_{2,i}(t|\theta)$, $i = 0, 1, ..., 2r-1$ are polynomials. In particular,

$$a_{2,0}(t|\theta) := \dot{Z}(\theta)'\Omega^{-1/2}a_1^{(2)}(t|\theta)\,\Omega^{-1/2}(\theta)\,\dot{Z}(\theta),$$

which is a first-order polynomial. Similarly, we find that $\partial^3 \log g_n(Z_n|\theta) / (\partial\theta\partial\theta'\partial\theta_i)$, $i = 1, ..., q$, can be expressed as a sum of ratios of polynomials of $T_n(\theta)$. Combining the above expressions, we see that $U_{n,k}$, $k = 1, 2, 3$, as defined in Lemma 3 are smooth functions of T_n which in turn implies that $-U_{n,1} + \frac{1}{\sqrt{n}}Q_1(U_{n,1}, U_{n,2}) + \frac{1}{n}Q_2(U_{n,1}, U_{n,2}, U_{n,3}) = f_n(T_n)$ for a smooth function f_n. It is easily checked that f_n satisfies the assumptions of Skovgaard (1981, Theorem 3.2) and so the distribution of $f_n(T_n)$ is well-approximated by a second-order Edgeworth expansion. This combined with the arguments of, for example, Taniguchi (1987, Lemma 4) in turn implies that the distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$ is well-approximated by a second-order Edgeworth expansion.

Next, we expand the BIL estimator around the MIL estimator. We here follow the strategy of Johnson (1970) who provides such an expansion in the case of the posterior mean based on the likelihood of a random sample. The arguments are almost identical to his given that, as shown above, we can control the errors when expanding the indirect likelihood. We therefore only sketch the proof and for notational simplicity only consider the case of a scalar parameter. First, for any given δ,

$$\pi\left(\hat{\theta}_{MIL} + n^{-1/2}\delta\right) = 1 + \sum_{i=1}^{4} n^{-i/2}\frac{1}{i!}\frac{\partial^i \pi\left(\hat{\theta}_{MIL}\right)}{\partial\theta^i}\delta^i + O_P\left(n^{-2}\right), \tag{17}$$

$$\log f_n\left(\hat{\theta}_{MIL} + n^{-1/2}\delta\right) = \log f_n\left(\hat{\theta}_{MIL}\right) + \sum_{i=2}^{4} n^{-i/2}\frac{1}{i!}l_{n,i}\left(\hat{\theta}_{MIL}\right)\delta^i + O_P\left(n^{-2}\right), \tag{18}$$

where $l_{n,i}(\theta) = \frac{1}{n}\partial \log f_n(\theta) / (\partial\theta)$. In particular, $l_{n,2}\left(\hat{\theta}_{MIL}\right) = -\mathcal{I}(\theta) + O_P\left(1/\sqrt{n}\right)$. Combining these two expansions, we obtain an expansion of the posterior density of

$\hat{\delta}_n = \sqrt{n}(\theta - \hat{\theta}_{MIL})$,

$$f_{\hat{\delta}_n}(\delta|Z_n) = \frac{\exp\left(\log f_n\left(\hat{\theta}_{MIL} + n^{-1/2}\delta\right) - \log f_n\left(\hat{\theta}_{MIL}\right)\right)\pi\left(\hat{\theta}_{MIL} + n^{-1/2}\delta\right)}{\int \exp\left(\log f_n\left(\hat{\theta}_{MIL} + n^{-1/2}\delta\right) - \log f_n\left(\hat{\theta}_{MIL}\right)\right)\pi\left(\hat{\theta}_{MIL} + n^{-1/2}\delta\right)d\delta};$$

see Johnson (1970) for details on this expansion. Next, using that $\theta = \hat{\theta}_{MIL} + \hat{\delta}_n/\sqrt{n}$, observe that the posterior mean can be expressed in terms of $f_n(\delta|Z_n)$,

$$\hat{\theta}_{BIL} = \int_\Theta \theta f_n(\theta|Z_n)\,d\theta = \hat{\theta}_{MIL} + \frac{1}{n}\int_\Theta \delta f_{\hat{\delta}_n}(\delta|Z_n)\,d\delta.$$

Substituting the expansion of $f_n(\delta|Z_n)$ into the last integral we obtain after some manipulations that

$$\hat{\theta}_{BIL} = \hat{\theta}_{MIL} + \left[\frac{1}{2}\mathcal{I}^{-1}(\theta) D^2 m \mathcal{I}^{-1}(\theta) + \mathcal{I}^{-1}(\theta)\frac{\dot{\pi}(\theta)}{\pi(\theta)}\right]\frac{1}{n} + \frac{1}{n^2}R_{BIL,n}, \tag{19}$$

where, for some ρ_n satisfying $\rho_n \to 0$ and $\rho_n\sqrt{n} \to \infty$, $P\left(\|R_{BIL,n}\| > \rho_n\sqrt{n}\right) = o\left(n^{-1}\right)$. In conclusion, the BIL estimator is equivalent to the MIL estimator up to a bias term of order $1/n$ and a stochastic term of order $1/n^2$. In particular, as claimed, the bias-adjusted BIL will enjoy the same properties as the bias-adjusted MIL, including second and third-order efficiency.

Finally, the CU-II estimator also falls within the framework of Lemma 3 with

$$m_n(\theta) = \dot{E}_\theta[Z_n]'\Omega^{-1}(\theta)\Delta_n(\theta) - \frac{1}{2\sqrt{n}}\Delta_n(\theta)'\frac{\partial\Omega^{-1}(\theta)}{\partial\theta}\Delta_n(\theta),$$

where we have redefined $T_n(\theta) = \sqrt{n}\Omega^{1/2}(\theta)(Z_n - E_\theta[Z_n])$ and $\Delta_n(\theta) = \Omega^{1/2}(\theta)T_n(\theta)$. In particular, compared to the MIL, $Z(\theta)$ has been replaced by $E_\theta[Z_n] \simeq Z(\theta) + B_Z(\theta)/n$. Otherwise, we can follow the same arguments as for MIL to obtain that

$$Q_1(U_{n,1}(\theta), U_{n,2}(\theta)) = n(Dm(\theta))^{-2}\left\{m_n(\theta)[Dm_n(\theta) - Dm(\theta)] + \frac{1}{2}D^2m(\theta)\right\},$$

where $Dm(\theta)$ and $D^2m(\theta)$ are the same as for the MIL. It holds that

$$\sqrt{n}m_n(\theta) = \dot{E}_\theta[Z_n]'\Omega^{-1/2}(\theta)T_n(\theta) + O_P\left(1/\sqrt{n}\right) = \dot{Z}(\theta)'\Omega^{-1/2}(\theta)T_n(\theta) + O_P\left(1/\sqrt{n}\right),$$

and

$$\begin{aligned}\sqrt{n}[Dm_n(\theta) - Dm(\theta)] &= \left[\ddot{E}_\theta[Z_n]'\Omega^{-1/2}(\theta) + 2\dot{E}_\theta[Z_n]'\frac{\partial\Omega^{-1}(\theta)}{\partial\theta}\Omega^{1/2}(\theta)\right]T_n(\theta) + O_P\left(1/\sqrt{n}\right)\\ &= \left[\ddot{Z}(\theta)'\Omega^{-1/2}(\theta) + 2\dot{Z}(\theta)'\frac{\partial\Omega^{-1}(\theta)}{\partial\theta}\Omega^{1/2}(\theta)\right]T_n(\theta) + O_P\left(1/\sqrt{n}\right),\end{aligned}$$

where we have used that $\dot{E}_\theta[Z_n] = \dot{Z}(\theta) + O(1/n)$ and similar for the second-order derivative. Thus, the CU-II estimator can also be expressed as a polynomial of T_n and so, by the same arguments as for the MIL estimator, its distribution satisfy an Edgeworth expansion. □

Proof. [Proposition 3] Observe that $\hat{f}_n(Z_n|\theta) = \sqrt{n/|\Omega(\theta)|}\hat{f}_{T_n}(T_n(Z_n|\theta)|\theta)$, where $T(Z_n|\theta) := \sqrt{n}\Omega^{-1/2}(\theta)(Z_n - Z(\theta))$ and, with $T_n^s(\theta) = \sqrt{n}\Omega^{-1/2}(\theta)(Z_n^s(\theta) - Z(\theta))$ and $B = \Omega^{-1/2}(\theta)\sqrt{n}h$, $\hat{f}_{T_n}(t|\theta) = \frac{1}{S}\sum_{s=1}^S K_B(T_n^s(\theta) - t)$. Given that $f_{T_n}(t|\theta)$ is suitably bounded uniformly in n

and θ, we can employ the results of Kristensen (2009) to obtain that

$$\sup_{n\geq 1}\sup_{\theta\in\Theta}\left|\hat{f}_{T_n}\left(T_n\left(Z_n|\theta\right)|\theta\right)-f_{T_n}\left(T_n\left(Z_n|\theta\right)|\theta\right)\right|=O_P\left(nh^2\right)+O_P\left(\sqrt{\log S/\left(S\left(\sqrt{n}h\right)^d\right)}\right).$$

This in turn implies that

$$\begin{aligned}\mathbb{I}\left\{\hat{f}_{n,S}(Z_n|\theta)>a\right\} &= \mathbb{I}\left\{\hat{f}_{T_n}\left(T_n\left(Z_n|\theta\right)|\theta\right)>a\left|\Omega\left(\theta\right)\right|/\sqrt{n}\right\}\\ &\simeq \mathbb{I}\left\{f_{T_n}\left(T_n\left(Z_n|\theta\right)|\theta\right)>a\left|\Omega\left(\theta\right)\right|/\sqrt{n}\right\}\\ &= \mathbb{I}\left\{f_n(Z_n|\theta)>a\right\}.\end{aligned}$$

Thus, by the mean value theorem, uniformly over $n\geq 1$ and $\theta\in\Theta$,

$$\begin{aligned}\hat{\tau}_a\left(Z_n|\theta\right)\frac{1}{n}\left|\log\hat{f}_n(Z_n|\theta)-\log f_n(Z_n|\theta)\right| &\simeq \frac{\mathbb{I}\left\{f_{T_n}\left(T_n\left(Z_n|\theta\right)|\theta\right)>a\left|\Omega\left(\theta\right)\right|/\sqrt{n}\right\}}{nf_{T_n}\left(T_n\left(Z_n|\theta\right)|\theta\right)}\left|\hat{f}_{T_n}\left(T_n\left(Z_n|\theta\right)|\theta\right)-f_{T_n}\left(T_n\left(Z_n|\theta\right)|\theta\right)\right|\\ &\leq \frac{C}{a\sqrt{n}}\left|\hat{f}_{T_n}\left(T_n\left(Z_n|\theta\right)|\theta\right)-f_{T_n}\left(T_n\left(Z_n|\theta\right)|\theta\right)\right|\\ &= O_P\left(a^{-1}n^{-1/2}h^2\right)+O_P\left(a^{-1}\sqrt{\log S/\left(Sn^{d+1}h^d\right)}\right).\end{aligned}$$

Finally, using that $P\left(\left\|Z_n-Z\left(\theta_0\right)\right\|^2/n<w\right)\to F\left(w\right)$ uniformly in w, where F is the cdf of a scaled χ^2_d-distribution,

$$\begin{aligned}P\left(\sup_{\theta\in\Theta}f_n(Z_n|\theta)>a_n\right) &\simeq P\left(\sup_{\theta\in\Theta}\exp\left(-\frac{1}{2}\left(Z_n-Z\left(\theta\right)\right)'\Omega\left(\theta\right)\left(Z_n-Z\left(\theta\right)\right)\right)>ca_n/\sqrt{n}\right)\\ &= P\left(\left\|Z_n-Z\left(\theta_0\right)\right\|^2/n<\frac{log\left(n/a_n\right)+c}{n}\right)\\ &\simeq F\left(\frac{log\left(n/a_n\right)}{n}\right)\simeq\exp\left(-\frac{log\left(n/a_n\right)}{2n}\right)\\ &= \exp\left(log\left(\left(n/a_n\right)^{-1/(2n)}\right)\right)=\left(n/a_n\right)^{-1/(2n)},\end{aligned}$$

for some constant $c>0$. With $q>0$ given in Assumption 4, choose $p>0$ such that $1/p+1/q=1$. We then obtain

$$\begin{aligned}&\frac{1}{n}E\left[\sup_{\theta\in\Theta}\left|\hat{\tau}_a\left(Z_n|\theta\right)-1\right|\left|\log f_n(Z_n|\theta)\right|\right]\\ \leq\ &\frac{1}{n}P\left(\sup_{\theta\in\Theta}f_n(Z_n|\theta)>2a_n\right)^{1/p}E\left[\sup_{\theta\in\Theta}\left|\log f_n(Z_n|\theta)\right|^q\right]^{1/q}\\ \simeq\ &\frac{1}{n}\left(n/a_n\right)^{-1/(2np)}.\end{aligned}$$

In total, uniformly over $\theta\in\Theta$,

$$\begin{aligned}&\frac{1}{n}\left|\hat{\tau}_a\left(Z_n|\theta\right)\log\hat{f}_n(Z_n|\theta)-\log f_n(Z_n|\theta)\right|\\ =\ &O_P\left(a^{-1}n^{-1/2}h^2\right)+O_P\left(a^{-1}\sqrt{\log S/\left(Sn^{d+1}h^d\right)}\right)+O_P\left(\frac{1}{n}\left(n/a_n\right)^{-1/(2np)}\right).\end{aligned}$$

The result now follows from Lemma 4.

By standard arguments for kernel regression estimators, see Li and Racine (2007, Sec. 2.1),

$$\hat{\theta}_{SBIL} \simeq \hat{\theta}_{BIL} + h^2 B_{BIL}(Z_n) + \sqrt{\frac{\|V_{BIL}(Z_n)\|}{Sh^d}},$$

where, with $f_n(Z_n) = \int_\Theta f_n(Z_n|\theta)\,\pi(\theta)\,d\theta$,

$$B_{BIL}(Z_n) = \frac{1}{2f_n(Z_n)} \sum_{i,j=1}^{d} \left\{ \frac{\partial^2 E_\theta[\theta|Z_n]}{\partial Z_{n,i}\partial Z_{n,j}} f_n(Z_n) + 2\frac{\partial E_\theta[\theta|Z_n]}{\partial Z_{n,i}} \frac{\partial f_n(Z_n)}{\partial Z_{n,j}} \right\} = O_P(n),$$

$$V_{BIL}(Z_n) = \int K^2(z)\,dz \times \frac{Var_\theta[\theta|Z_n]}{f_n(Z_n)} = O_P\left(n^{d/2}\right).$$

□

Appendix B: Lemmas

Lemma 1. *Suppose that the remainder term $R_n(t|\theta)$ in eq. (6) satisfies for some $c \geq 0$ and $q \geq 1$,*

$$\sup_\theta |R_n(t|\theta)| \leq \frac{c}{n^{(r+1)/2}} \frac{1}{1+|t|^q}.$$

Then Assumption 1(i) and (iii) are satisfied. If, furthermore, we are allowed to differentiate on both sides of eq. (6) w.r.t. θ, and the derivatives of $R_n(t|\theta)$ satisfies the above bound, then 1(ii) holds.

Proof. Observe that, using change of variable, $t = \sqrt{n}\Omega^{-1/2}(\theta_0)(z - Z(\theta_0)) \Leftrightarrow z = \Omega^{1/2}(\theta_0)\,t/\sqrt{n} + Z(\theta_0)$ such that

$$T(z|\theta) := \sqrt{n}\Omega^{-1/2}(\theta)\left(\Omega^{1/2}(\theta_0)\,z/\sqrt{n} + Z(\theta_0) - Z(\theta)\right) = V(\theta)\,z + \sqrt{n}\delta(\theta)$$

where $V(\theta) = \Omega^{-1/2}(\theta)\,\Omega^{1/2}(\theta_0)$ and $\delta(\theta) = \Omega^{-1/2}(\theta)\,[Z(\theta_0) - Z(\theta)]$. Using this transformation,

$$\begin{aligned}
& E\left[\sup_{\theta\in\Theta} |LR(\theta)|\right] \\
(20) = & \frac{1}{n}\int f_n(z|\theta_0) \sup_\theta |\log f_n(z|\theta) - \log f_n^*(z|\theta)|\,dz \\
= & \frac{1}{n}\int f_{T_n}(t|\theta_0) \sup_\theta \left|\log f_{T_n}\left(V(\theta)\,t + \sqrt{n}\delta(\theta)\,|\theta\right) - \log f_{T_n}^*\left(V(\theta)\,t + \sqrt{n}\delta(\theta)\,|\theta\right)\right| dt \\
(21) \leq & \frac{2}{n}\int \phi(t) \sup_\theta \left|\log\left(1 + 2\frac{R_n\left(V(\theta)\,t + \sqrt{n}\delta(\theta)\,|\theta\right)}{\phi\left(V(\theta)\,t + \sqrt{n}\delta(\theta)\,|\theta\right)}\right)\right| dt,
\end{aligned}$$

where we have used that $\phi(t)/2 \leq f_{T_n}(t|\theta) \leq 2\phi(t)$ for n large enough. Observe that for some $c_1, c_2 > 0$,

$$\frac{1}{n}\sup_{\theta\in\Theta}\left|\log\left(1 + 2\frac{R_n\left(V(\theta)\,t + \sqrt{n}\delta(\theta)\,|\theta\right)}{\phi\left(V(\theta)\,t + \sqrt{n}\delta(\theta)\,|\theta\right)}\right)\right| \leq \log\left(1 + \frac{c}{n^{(r+1)/2n}} \frac{\exp\left(c_1 + c_2|t|_1^2/n\right)}{1 + c_1 n^q + c_2|t|^q}\right) \to 0.$$

Since the function on the right hand side of the above inequality is also integrable w.r.t. $\phi(t)$, it follows by the dominated convergence theorem that the bound in eq. (21) goes to

zero. That is, Assumption 1(i) is satisfied. For Assumption 1(iii), define $\theta_n = \theta_0 + \tau/\sqrt{n}$ and use the mean-value theorem to obtain $\delta(\theta_n) \simeq \dot{Z}(\theta_0)\tau/\sqrt{n}$. Substituting this into the integral bound in eq. (21), we obtain

$$\begin{aligned} E\left[\sup_{\theta_n}|LR(\theta_n)|\right] &\leq \frac{2}{n}\int \phi(t)\sup_{\tau}\left|\log\left(1+2\frac{R_n\left(V(\theta_n)t+\dot{Z}(\theta_0)\tau|\theta_0+\tau/\sqrt{n}\right)}{\phi\left(V(\theta_n)t+\dot{Z}(\theta_0)\tau|\theta_0+\tau/\sqrt{n}\right)}\right)\right|dt \\ &\leq \frac{2}{n}\int \phi(t)\left|\log\left(1+\frac{c}{n^{(r+1)/2}}\frac{\exp\left(c_1+c_2|t|_1^2\right)}{1+c_1+c_2|t|^q}\right)\right|dt, \end{aligned}$$

For any t,

$$\log\left(1+\frac{c}{n^{(r+1)/2}}\frac{\exp\left(c_1+c_2|t|_1^2\right)}{1+c_1+c_2|t|^q}\right) = O\left(n^{(r+1)/2}\right),$$

and so Assumption 1(iii) holds by the dominated convergence theorem. Assumption 1(ii) is shown along the same lines, and so the proof is left out. □

Lemma 2. *Under the assumptions of Proposition 2, the leading bias terms of the MIL, BIL and CU estimators are:*

$$E_\theta[\hat{\theta}_{MIL}] \simeq \theta + \tfrac{1}{n}B_I(\theta) + \frac{1}{n}B_{MIL}(\theta), \qquad E_\theta[\hat{\theta}_{BIL}] \simeq \theta + \frac{1}{n}B_I(\theta) + \frac{1}{n}B_{MIL}(\theta) + \frac{1}{n}B_{BIL}(\theta),$$

$$E_\theta[\hat{\theta}_{CU}] \simeq \theta + \tfrac{1}{n}B_I(\theta) + \frac{1}{n}B_{CU}(\theta),$$

where, with $a_1^{(2)}(t|\theta) = \partial^2 a_1(t|\theta)/(\partial t \partial t)$,

$$B_I(\theta) = \frac{1}{2}\mathcal{I}^{-2}(\theta)\dot{Z}(\theta)'\left\{\Omega^{-1}(\theta)\ddot{Z}(\theta) - \frac{\partial\Omega^{-1}(\theta)}{\partial\theta}\dot{Z}(\theta)\right\} \qquad B_{CU}(\theta) = 2tr\left\{\frac{\partial\Omega^{-1}(\theta)}{\partial\theta}\Omega(\theta)\right\},$$

$$B_{MIL}(\theta) = \mathcal{I}^{-2}(\theta)\dot{Z}(\theta)'\Omega^{-1/2}(\theta)E_\theta\left[T_n(\theta)\dot{Z}(\theta)'\Omega^{-1/2}(\theta)a_1^{(2)}(T_n(\theta)|\theta)\right]\Omega^{-1/2}(\theta)\dot{Z}(\theta),$$

$$B_{BIL}(\theta) = \frac{3}{2}\mathcal{I}^{-2}(\theta)\dot{Z}(\theta)'\left\{\Omega^{-1}(\theta)\ddot{Z}(\theta) + \frac{\partial\Omega^{-1}(\theta)}{\partial\theta}\dot{Z}(\theta)\right\} + \mathcal{I}^{-1}(\theta)\frac{\dot{\pi}(\theta)}{\pi(\theta)}.$$

Proof. The leading bias of $\hat{\theta}_{MIL}$ could in principle be found by writing up the Edgeworth expansion of $\sqrt{n}(\hat{\theta}-\theta_0)$ explicitly. We here instead directly analyze the mean of $Q_1(U_{n,1}, U_{n,2})/n$ as defined in the proof of Proposition 2. First,

$$Q_1(U_{n,1}(\theta), U_{n,2}(\theta)) = n(Dm(\theta))^{-2}\left\{m_n(\theta)[Dm_n(\theta) - Dm(\theta)] + \frac{1}{2}D^2m(\theta)\right\},$$

where

$$\begin{aligned} \sqrt{n}m_n(\theta) &= \dot{Z}(\theta)'\Omega^{-1/2}(\theta)T_n(\theta) + O_P\left(1/\sqrt{n}\right), \\ \sqrt{n}\left[Dm_n(\theta) - Dm(\theta)\right] &= \left[\ddot{Z}(\theta)'\Omega^{-1/2}(\theta) + 2\dot{Z}(\theta)'\frac{\partial\Omega^{-1}(\theta)}{\partial\theta}\Omega^{1/2}(\theta)\right]T_n(\theta) \\ &\quad + \dot{Z}(\theta)'\Omega^{-1/2}(\theta)a_1^{(2)}(T_n(\theta)|\theta)\Omega^{-1/2}(\theta)\dot{Z}(\theta) + O_P\left(1/\sqrt{n}\right). \end{aligned}$$

Substituting the leading terms into the above expression for $Q_1\left(U_{n,1}, U_{n,2}\right)$ yields

$$\begin{aligned} E_\theta\left[Q_1\left(U_{n,1}\left(\theta\right), U_{n,2}\left(\theta\right)\right)\right] &\simeq J^{-2}\left(\theta\right)\left\{\dot{Z}\left(\theta\right)'\Omega^{-1}\left(\theta\right)\ddot{Z}\left(\theta\right)+2\dot{Z}\left(\theta\right)'\frac{\partial\Omega^{-1}\left(\theta\right)}{\partial\theta}\dot{Z}\left(\theta\right)\right\}+B_{MIL}\left(\theta\right) \\ &\quad -\frac{3}{2}J^{-2}\left(\theta\right)\left\{\dot{Z}\left(\theta\right)'\Omega^{-1}\left(\theta\right)\ddot{Z}\left(\theta\right)+\dot{Z}\left(\theta\right)'\frac{\partial\Omega^{-1}\left(\theta\right)}{\partial\theta}\dot{Z}\left(\theta\right)\right\} \\ &= \frac{1}{2}J^{-2}\left(\theta\right)\dot{Z}\left(\theta\right)'\left\{\Omega^{-1}\left(\theta\right)\ddot{Z}\left(\theta\right)-\frac{\partial\Omega^{-1}\left(\theta\right)}{\partial\theta}\dot{Z}\left(\theta\right)\right\}+B_{MIL}\left(\theta\right), \end{aligned}$$

where

$$B_{MIL}\left(\theta\right):=J^{-2}\left(\theta\right)\dot{Z}\left(\theta\right)'\Omega^{-1/2}\left(\theta\right)E\left[T_n\left(\theta\right)\dot{Z}\left(\theta\right)'\Omega^{-1/2}\left(\theta\right)a_1^{(2)}\left(T_n\left(\theta\right)|\theta\right)\right]\Omega^{-1/2}\left(\theta\right)\dot{Z}\left(\theta\right)$$

Thus, using that $E_\theta\left[U_{n,1}\left(\theta\right)\right]=0$,

$$\begin{aligned} E_\theta[\hat{\theta}_{MIL}]-\theta &\simeq \frac{1}{n}E_\theta\left[Q_1\left(U_{n,1}\left(\theta\right), U_{n,2}\left(\theta\right)\right)\right] \\ &\simeq \frac{1}{2n}J^{-2}\left(\theta\right)\left\{\dot{Z}\left(\theta\right)'\Omega^{-1}\left(\theta\right)\ddot{Z}\left(\theta\right)-\dot{Z}\left(\theta\right)'\frac{\partial\Omega^{-1}\left(\theta\right)}{\partial\theta}\dot{Z}\left(\theta\right)\right\}+\frac{1}{n}B_{MIL}\left(\theta\right). \end{aligned}$$

The leading bias of the BIL estimators is obtained by combining eq. (19) with the expression of the bias for the MIL estimator. Finally, observe that the first-order condition defining the CU estimator satisfies $E_\theta\left[m_n\left(\theta\right)\right]=\frac{1}{2\sqrt{n}}tr\left\{\frac{\partial\Omega^{-1}(\theta)}{\partial\theta}\Omega\left(\theta\right)\right\}$, and so

$$\begin{aligned} E_\theta[\hat{\theta}_{CU}]-\theta &\simeq \frac{1}{\sqrt{n}}E\left[m_n\left(\theta\right)\right]+\frac{1}{n}E\left[Q_1\left(U_{n,1}\left(\theta\right), U_{n,2}\left(\theta\right)\right)\right] \\ &\simeq \frac{1}{2}J^{-2}\left(\theta\right)\left\{\dot{Z}\left(\theta\right)'\Omega^{-1}\left(\theta\right)\ddot{Z}\left(\theta\right)-\dot{Z}\left(\theta\right)'\frac{\partial\Omega^{-1}\left(\theta\right)}{\partial\theta}\dot{Z}\left(\theta\right)\right\}+\frac{1}{n}B_{CU}\left(\theta\right). \end{aligned}$$

□

Consider some $\sqrt{n}$-consistent estimator $\hat{\theta}\in\mathbb{R}^q$, $\sqrt{n}\left(\hat{\theta}-\theta_0\right)=O_P\left(1\right)$, characterized as the root of a random function $m_n\left(\theta\right)\in\mathbb{R}^q$:

$$0=m_n\left(\hat{\theta}\right).$$

We wish to analyze higher-order properties of $\hat{\theta}$. For that purpose we introduce the following sequences:

$$U_{n,1}=-\sqrt{n}Dm^{-1}m_n\left(\theta_0\right),\ U_{n,2}=\sqrt{n}\left\{\frac{\partial m_n\left(\theta_0\right)}{\partial\theta}-Dm\right\},\ U_{n,3,i}=\sqrt{n}\left\{\frac{\partial^2 m_n\left(\theta_0\right)}{\partial\theta\partial\theta_i}-D^2m_i\right\},$$

for $i=1,...,q$, where $Dm\in\mathbb{R}^{q\times q}$, $D^2m_i\in\mathbb{R}^{q\times q}$ and $D^3m_{ij}\in\mathbb{R}^{q\times q}$; in the leading case, these will be moments.

Lemma 3. *Suppose that the stochastic function $\theta\mapsto m_n\left(\theta\right)$ is three times differentiable with its derivatives satisfying $U_{n,k}=O_P\left(1\right)$ for $k=1,2,3$ and the matrix $Dm\in\mathbb{R}^{q\times q}$ is non-singular. Moreover, for some sequence $D_n=O_P\left(1\right)$,*

$$\left|\frac{\partial^3 m_n\left(\theta\right)}{\partial\theta\partial\theta_i\partial\theta_j}-\frac{\partial^3 m_n\left(\theta_0\right)}{\partial\theta\partial\theta_i\partial\theta_j}\right|\leq D_n\left|\theta-\theta_0\right|,$$

in a neighbourhood of θ_0. Then the estimator satisfies

$$\sqrt{n}(\hat{\theta}-\theta_0) = U_{n,1} + \frac{1}{\sqrt{n}}Q_1\left(U_{n,1},U_{n,2}\right) + \frac{1}{n}Q_2\left(U_{n,1},U_{n,2},U_{n,3}\right) + \frac{1}{n^{3/2}}R_n,$$

where $R_n = O_P(1)$ and

$$Q_1\left(U_{n,1},U_{n,2}\right) = -Dm^{-1}\left\{U_{n,2}U_{n,1} + \frac{1}{2}\sum_{i=1}^{q}U_{n,1}D^2m_iU_{n,1}\right\},$$

$$\begin{aligned} & Q_2\left(U_{n,1},U_{n,2},U_{n,3}\right) \\ = \; & -Dm^{-1}U_{n,2}Q_1\left(U_{n,1},U_{n,2}\right) - \frac{1}{6}Dm^{-1}\sum_{i,j=1}^{q}U_{n,1,i}U_{n,1,j}Dm_{ij}U_{n,1} \\ & -Dm^{-1}\frac{1}{2}\sum_{i=1}^{q}\left\{U_{n,1,i}D^2m_iQ_1\left(U_{n,1},U_{n,2}\right) + Q_{1,i}\left(U_{n,1},U_{n,2}\right)D^2m_iU_{n,1} + U_{n,1,i}U_{n,3,i}U_{n,1}\right\}. \end{aligned}$$

If, in addition, for each $\alpha > 0$ there exists $d > 0$ such that

$$P\left(\|m_n(\theta_0)\| > dn^{\alpha}\right) = o\left(n^{-1}\right),\;\; P\left(\|B_{n,i}\| > dn^{\alpha}\right) = o\left(n^{-1}\right),\;\; P\left(\left\|C_{n,ij}\right\| > dn^{\alpha}\right) = o\left(n^{-1}\right),$$

$$P\left(sup_{\theta}\left\|\frac{\partial^3 m_n(\theta)}{\partial\theta\partial\theta_i\partial\theta_j} - \frac{\partial^3 m_n(\theta_0)}{\partial\theta\partial\theta_i\partial\theta_j}\right\| > dn^{\alpha}\left|\theta-\theta_0\right|\right) = o\left(n^{-1}\right),$$

then, for some ρ_n satisfying $\rho_n \to 0$ and $\rho_n\sqrt{n} \to \infty$, $P\left(\|R_n\| > \rho_n\sqrt{n}\right) = o\left(n^{-1}\right)$.

Proof. We proceed as in Newey and Smith (2004): First, by a third order Taylor expansion,

$$\begin{aligned} 0 \;=\; & m_n(\theta_0) + \frac{\partial m_n(\theta_0)}{\partial\theta}\left(\hat{\theta}-\theta_0\right) + \frac{1}{2}\sum_{i=1}^{q}\left(\hat{\theta}_i-\theta_{0,i}\right)\frac{\partial^2 m_n(\theta_0)}{\partial\theta\partial\theta_i}\left(\hat{\theta}-\theta_0\right) \\ & + \frac{1}{6}\sum_{i,j=1}^{q}\left(\hat{\theta}_i-\theta_{0,i}\right)\left(\hat{\theta}_j-\theta_{0,j}\right)\frac{\partial^3 m_n(\bar{\theta})}{\partial\theta\partial\theta_i\partial\theta_j}\left(\hat{\theta}-\theta_0\right), \end{aligned}$$

where $\bar{\theta}$ lies on the line between θ_0 and $\hat{\theta}$. Since the third order derivative satisfies

$$\left|\frac{\partial^3 m_n(\bar{\theta})}{\partial\theta\partial\theta_i\partial\theta_j} - D^3m_{ij}\right| \leq \left|\frac{\partial^3 m_n(\bar{\theta})}{\partial\theta\partial\theta_i\partial\theta_j} - \frac{\partial^3 m_n(\theta_0)}{\partial\theta\partial\theta_i\partial\theta_j}\right| + O_P\left(1/\sqrt{n}\right) = O_P\left(1/\sqrt{n}\right),$$

we obtain

$$\begin{aligned} 0 \;=\; & \sqrt{n}m_n(\theta_0) + \sqrt{n}Dm\left(\hat{\theta}-\theta_0\right) + U_{n,2}\left(\hat{\theta}-\theta_0\right) \\ & + \frac{\sqrt{n}}{2}\sum_{i=1}^{q}\left(\hat{\theta}_i-\theta_{0,i}\right)D^2m_i\left(\hat{\theta}-\theta_0\right) + \frac{1}{2}\sum_{i=1}^{q}\left(\hat{\theta}_i-\theta_{0,i}\right)U_{n,3,i}\left(\hat{\theta}-\theta_0\right) \\ & + \frac{\sqrt{n}}{6}\sum_{i,j=1}^{q}\left(\hat{\theta}_i-\theta_{0,i}\right)\left(\hat{\theta}_j-\theta_{0,j}\right)D^3m_{ij}\left(\hat{\theta}-\theta_0\right) + O_P\left(1/n^{3/2}\right). \end{aligned}$$

Next, using that $U_{n,1} = O_P(1)$ and $U_{n,2} = O_P(1)$, the result now follows by the same arguments as in Newey and Smith (2004, Proof of Lemma A4). The second part stating that $P\left(\|R_n\| > \rho_n\sqrt{n}\right) = o\left(n^{-1}\right)$ follows by standard arguments for Edgeworth expansions of estimators; see e.g. Taniguchi (1987, p. 8-10). □

Lemma 4. *Let* $\hat{\theta} = \arg\min_{\theta\in\Theta} L_n(\theta)$ *and* $\tilde{\theta} = \arg\min_{\theta\in\Theta} Q_n(\theta)$ *satisfy: (i)* $\hat{\theta} \to^P \theta_0$ *and* $\tilde{\theta} \to^P \theta_0$*; (ii)* $L_n(\theta)$ *is twice differentiable with* $\sup_{\|\theta-\theta_0\|<\delta} \left\|\partial^2 L_n(\theta) / (\partial\theta\partial\theta') - H(\theta)\right\| \to^P 0$*; (iii)* $H(\theta)$ *is continuous with* $H(\theta_0) > 0$*. Then*

$$\left\|\hat{\theta} - \tilde{\theta}\right\| = O_P\left(\sqrt{\sup_{\|\theta-\theta_0\|<\delta} |L_n(\theta) - Q_n(\theta)|}\right).$$

Let $\hat{\theta}$ *and* $\tilde{\theta}$ *solve* $S_n(\hat{\theta}) = 0$ *and* $T_n(\tilde{\theta}) = 0$*, respectively, such that: (i)* $\hat{\theta} \to^P \theta_0$ *and* $\tilde{\theta} \to^P \theta_0$*; (ii)* $S_n(\theta)$ *is differentiable with* $\sup_{\|\theta-\theta_0\|<\delta} \|\partial S_n(\theta)/\partial\theta - H(\theta)\| \to^P 0$*; (iii)* $H(\theta)$ *is continuous with* $H(\theta_0) > 0$*. Then*

$$\left\|\hat{\theta} - \tilde{\theta}\right\| = O_P\left(\sup_{\|\theta-\theta_0\|<\delta} |S_n(\theta) - T_n(\theta)|\right).$$

APPENDIX C: IMPLEMENTATION OF SMIL AND SBIL

To compute both SMIL and SBIL, we use the k nearest neighbors approach (see Li and Racine, 2007, Chapter 14). The number of neighbors used for nonparametric fits is set to $k = a \times S^{0.25}$, rounded down to the nearest integer, where S is the number of simulations drawn from the pseudo-prior. For the auction model, which is very fast to simulate, we use $S = 2 \times 10^6$, while for the DSGE model we use $S = 10^6$. For the auction model, which has a fairly low dimensional auxiliary statistic, we use $a = 1.5$, while for the DSGE model, which uses a higher dimensional auxiliary statistic, we set $a = 1$. More careful choice using methods such as cross validation might improve the results somewhat, but we do not explore this possibility in this paper. For all applications the pseudo-prior $\pi(\theta)$ is a uniform distribution over the parameter space Θ, so the only remaining issue is specifying the bounds of parameter space.

We compute the SBIL estimator as the posterior mean. Similar results are obtained if the posterior median is used instead - these are not reported to save space. To compute the SMIL estimator, we maximize the joint density of (Z_n, θ) with respect to θ, where the joint density is estimated using a $k-$nearest neighbors nonparametric density fit computed using the same S simulations (θ^s, Z_n^s), $s = 1, ...S$ that are used to compute the SBIL estimator. Given that our prior is uniform, the joint density of (Z_n, θ) is maximized at the same value of θ as is the likelihood function. We use this strategy to take advantage of the simulations that have already been done for SBIL, rather than using the method outlined in equation 13, which would require new simulations at each trial value of θ. For a simple nearest neighbors density fit (Li and Racine, 2007, equation 14.2), maximizing the density is equivalent to minimizing the distance of the k-th neighbor to the observed Z_n. The objective function is not convex, so a global minimization algorithm (simulated annealing) is used. This is quite time-consuming compared to SBIL, which can be computed in less than a minute after the simulations have been done. In all cases we compute 5000 Monte Carlo replications of the estimators. All software used to compute the results is free, and complete code and all software required to replicate all results reported in this paper is available from the authors.

DSGE Model. The implementation of the SBIL estimator for this model proceeds as follows: Given a draw θ^s from the parameter space, first, the model is solved using Dynare (Adjemian et al., 2011), using a third order perturbation about the steady state. Then a simulation of length $n + 100$ is done, initialized at the steady state. We drop the initial 100 observations, retaining the last n observations, where n is either 40 or 160. The selection of observable variables is in line with much empirical work (e.g., Smets and Wouters, 2007, also see Guerrón-Quintana, 2010, for discussion). Given the sample of observed variables, we compute the auxiliary statistic Z_n^s.

Our pseudo-prior $\pi(\theta)$ is chosen as a uniform distribution over the hypercube defined by the bounds of the parameter space, which are found in Table 2. Following Ruge-Murcia (2012), rather than set a prior for ψ and estimate this parameter directly, we instead treat steady state hours $\bar{n}$ as a parameter to estimate, along with the other parameters, excepting ψ. Then ψ is recovered using the equation $\psi = \bar{c}^{-\gamma} (1 - \alpha) \bar{k}^{\alpha} \bar{n}^{-\alpha}$, where

overbars indicate the steady state value of a variable (the steady state solution is given in the Dynare code for the model). The advantage of this is that it is comparatively easy to set priors on $\bar{n}$, which guarantees that we will not be wasting time generating many unrealistic solutions where the average number of hours worked is far away from the assumed true value, which is 1/3 of the time endowment (8 hours a day), for both designs. The chosen limits cause the pseudo-prior means to be biased for the true parameter values. The chosen limits are intended to be broad, in comparison to the fairly strongly informative priors that are often used when estimating DSGE models (Fernández-Villaverde, 2010, discusses use of strongly informative priors).

The elements of the auxiliary statistic are chosen with an eye to their ability to identify the parameters of the model, and include variable means, standard deviations, autocorrelations, and some statistics resulting from regressions between variables. We implement two versions of SBIL, a first that uses all statistics to estimate all parameters, and a targeted version, where specific statistics are used for each parameter, to reduce the dimension when doing the nonparametric regression. The first version exactly follows the theoretical presentation. The targeted version may be thought of as a collection of SBIL estimators, each of which follows the theory individually, and from which we select out estimators of the parameters that are targeted by the respective vector of statistics. The advantage of the targeting is that it allows for reduction of the dimension of the conditioning information in the nonparmametric estimation of $E(\theta_j|Z_n)$, as Z_n is chosen specifically for each parameter θ_j, $j = 1, 2, ..., k$. This results in more accurate nonparametric estimation, for a given number of simulations, S.

We now discuss the targeted statistics, as the overall statistic is simply the union of the targeted statistics. For each of the parameters α, β, δ and $\bar{n}$, the targeted auxiliary statistics are the sample means of the five observable variables. Some of the other auxiliary statistics are more complicated. For example, the model implies that $w = \psi\eta c^{\gamma}$, so $\log w = \log\psi + \gamma \log c + \log\eta$, where $\log\eta$ follows an AR(1) process. Because w and c are observable, this equation can be estimated. We use a generalized instrumental variables estimator (GIV), using the lags of the logarithms of the observable variables as instruments. The GIV estimation results give a statistic $\hat{\gamma}$ which is used as one of the targeted auxiliary statistics for the parameter γ. Other statistics used for γ are sample correlations between consumption and the other observable variables. The residuals from the GIV estimation can be used to fit the regression $\widehat{\log\eta_t} = \rho_\eta \widehat{\log\eta_{t-1}} + \epsilon_t$, which leads to statistics that are informative for the parameters ρ_η and σ_η. For IL estimation of ρ_η, we supplement this statistic with the sample first order autocovariances of consumption with the full set of observable variables. The auxiliary statistics used for ρ_z and σ_z use a similar auxiliary regression, along with autocovariances between output and the other variables, in the case of ρ_z. The exact details of all of the auxiliary statistics used are given in the provided code, in the files `aux_stat.m`, which computes all auxiliary statistics from the simulated data, and in the body of the main file `DSGE_SBIL_Simple.m`, following the comment line "select statistics for each parameter".

Given that the SBIL estimator is simulation-based, experimentation can be done to identify what statistics are informative for each parameter. A systematic approach to

how to use experimentation to choose targeted statistics may be an interesting topic for further research.

APPENDIX D: TABLES AND FIGURES

TABLE 1. Auction model. Monte Carlo results (5000 replications).

	$\theta_0 = 0.5$		$\theta_1 = 0.5$	
Estimator	Bias	RMSE	Bias	RMSE
CU-II, $n = 20$	-0.025	0.102	0.042	0.181
SBIL, $n = 20$	-0.004	0.060	0.009	0.113
CU-II, $n = 80$	-0.006	0.052	0.013	0.101
SBIL, $n = 80$	0.001	0.036	-0.003	0.069
SMIL, $n = 80$	0.003	0.046	-0.006	0.088
CU-II, $n = 320$	.0.003	0.027	-0.004	0.055
SBIL, $n = 320$	0.002	0.022	-0.002	0.042

TABLE 2. DSGE models, support of uniform priors.

Parameter	Lower bound	Upper bound
α	0.2	0.4
β	0.9	0.999
δ	0.005	0.1
γ	0.0	3.0
ρ_z	0.0	0.999
σ_z	0.001	0.1
ρ_η	0.0	0.999
σ_η	0.001	0.1
$\bar{n}$	1/4	1/2

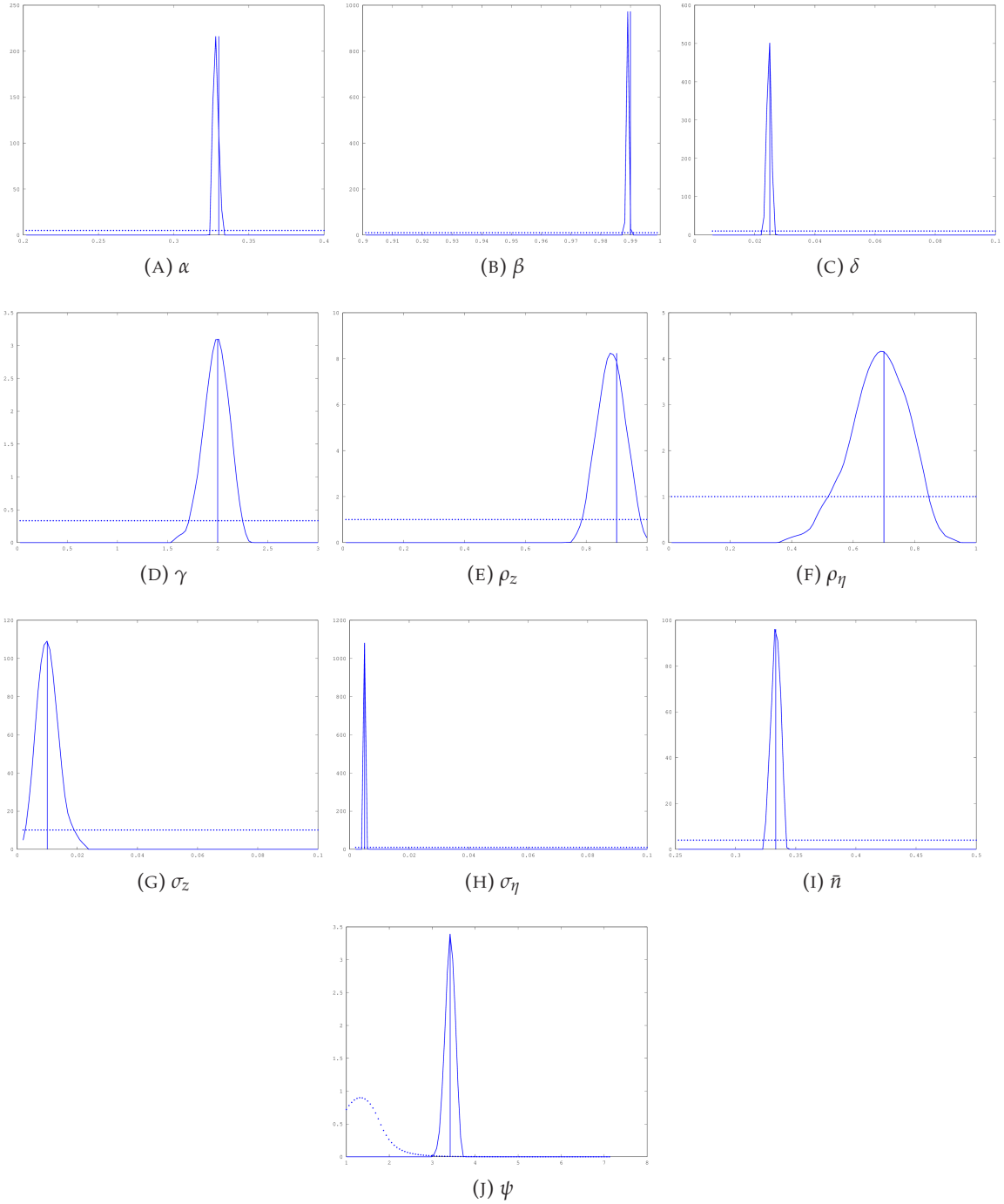
TABLE 3. DSGE model, first design. Monte Carlo results (5000 replications).

		Bias				RMSE			
		Standard IL		Targeted IL		Standard IL		Targeted IL	
Parameter	True value	40	160	40	160	40	160	40	160
α	0.330	0.003	0.014	-0.002	-0.002	0.013	0.0171	0.003	0.002
β	0.990	-0.009	-0.005	-0.001	-0.001	0.010	0.006	0.001	0.001
δ	0.025	0.004	0.005	0.000	-0.000	0.006	0.006	0.002	0.001
γ	2.000	-0.039	-0.091	-0.020	-0.015	0.264	0.199	0.198	0.096
ρ_z	0.900	-0.063	-0.033	-0.128	-0.018	0.086	0.044	0.173	0.038
σ_z	0.010	0.055	0.048	0.000	0.000	0.056	0.048	0.004	0.003
ρ_η	0.700	-0.060	-0.034	-0.058	-0.016	0.138	0.075	0.145	0.072
σ_η	0.005	0.027	0.016	0.001	0.000	0.030	0.017	0.007	0.000
$\bar{n}$	1/3	0.013	0.000	-0.002	0.000	0.021	0.008	0.008	0.003
ψ	3.417	-0.143	-0.074	0.045	-0.005	0.353	0.206	0.206	0.086

TABLE 4. DSGE model, second design. Monte Carlo results (5000 replications).

		Bias				RMSE			
		Standard IL		Targeted IL		Standard IL		Targeted IL	
Parameter	True value	40	160	40	160	40	160	40	160
α	0.250	0.014	0.012	-0.002	-0.004	0.020	0.014	0.007	0.005
β	0.970	-0.003	0.000	-0.001	-0.002	0.006	0.005	0.002	0.002
δ	0.040	0.003	0.004	-0.000	-0.001	0.007	0.008	0.002	0.001
γ	1.000	0.137	0.021	0.077	0.014	0.237	0.093	0.140	0.055
ρ_z	0.800	-0.080	-0.025	-0.084	-0.008	0.117	0.038	0.121	0.056
σ_z	0.020	0.042	0.037	0.002	-0.004	0.043	0.038	0.008	0.005
ρ_η	0.600	0.018	0.006	0.034	0.021	0.117	0.050	0.112	0.056
σ_η	0.010	0.018	0.015	0.002	0.001	0.021	0.016	0.003	0.001
$\bar{n}$	0.333	-0.001	-0.001	0.003	-0.002	0.009	0.005	0.006	0.002
ψ	2.619	0.319	0.069	0.156	0.039	0.565	0.240	0.312	0.127

FIGURE 1. DSGE model, first design, n=160. Pseudo-priors (dotted curve), true parameter values (vertical line), and Monte Carlo density of SBIL (solid curve)



(A) α (B) β (C) δ

(D) γ (E) ρ_z (F) ρ_η

(G) σ_z (H) σ_η (I) $\bar{n}$

(J) ψ

UNIVERSITAT AUTÒNOMA DE BARCELONA, BARCELONA GRADUATE SCHOOL OF ECONOMICS, AND MOVE

UNIVERSITY COLLEGE LONDON, CEMMAP (CENTRE FOR MICRODATA METHODS AND PRACTICE, IFS) AND CREATES (CENTER FOR RESEARCH IN ECONOMETRIC ANALYSIS OF TIME SERIES, UNIVERSITY OF AARHUS).