Indirect Likelihood Inference

Michael Creel
Dennis Kristensen

INDIRECT LIKELIHOOD INFERENCE

MICHAEL CREEL AND DENNIS KRISTENSEN

ABSTRACT. Given a sample from a fully specified parametric model, let Z_n be a given finite-dimensional statistic - for example, an initial estimator or a set of sample moments. We propose to (re-)estimate the parameters of the model by maximizing the likelihood of Z_n. We call this the maximum indirect likelihood (MIL) estimator. We also propose a computationally tractable Bayesian version of the estimator which we refer to as a Bayesian Indirect Likelihood (BIL) estimator. In most cases, the density of the statistic will be of unknown form, and we develop simulated versions of the MIL and BIL estimators. We show that the indirect likelihood estimators are consistent and asymptotically normally distributed, with the same asymptotic variance as that of the corresponding efficient two-step GMM estimator based on the same statistic. However, our likelihood-based estimators, by taking into account the full finite-sample distribution of the statistic, are higher order efficient relative to GMM-type estimators. Furthermore, in many cases they enjoy a bias reduction property similar to that of the indirect inference estimator. Monte Carlo results for a number of applications including dynamic and nonlinear panel data models, a structural auction model and two DSGE models show that the proposed estimators indeed have attractive finite sample properties.

Keywords: indirect inference; maximum-likelihood; simulation-based methods; bias correction; Bayesian estimation.

JEL codes: C13, C14, C15, C33.

1. INTRODUCTION

Suppose we have a fully specified and thus simulable model, indexed by a parameter $\theta \in \Theta \subset \mathbb{R}^k$. We have observed a sample $Y_n = (y_1, ..., y_n)$ generated at the unknown true parameter value θ_0 about which we wish to learn. A natural tool to this end is the likelihood function, $f(Y_n|\theta)$, and the associated maximum likelihood estimator (MLE), which has a number of attractive large sample optimality properties. However, the MLE is in some situations difficult to compute due to the complexity of the model, and it 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. Moreover, even if the MLE is easily computed, it may suffer from significant biases in finite samples with the resulting precision being rather poor, which complicates finite-sample inference. Well-known examples are the biases of least-squares estimators in autoregressive models (Andrews, 1993) and in dynamic and nonlinear panel data models (Hahn and Kuersteiner, 2002; Hahn and Newey, 2004).

To deal with the issue of computational complexity, researchers often resort to GMM-type methods where a statistic $Z_n = Z_n(Y_n)$ is used to draw inference regarding the parameter of interest. Suppose for example, that Z_n is a set of sample moments: Then a natural way to estimate parameters is to minimize the distance between sample and model-implied moments. When the form of the population moments are unknown, simulations may be used, and one obtains the simulated method of moments (SMM; McFadden, 1989; Duffie and Singleton, 1993). The indirect inference estimator (II; Gouriéroux, Monfort, Renault, 1993; Smith, 1993) proposes an alternative choice for Z_n, namely as an extremum estimator based on an auxiliary model. The efficient method of moments (EMM; Gallant and Tauchen, 1996) sets Z_n to be the score vector of an auxiliary model.

Similarly, there exist numerous methods designed to reduce biases in estimators such as bootstrap (Everaert and Pozzi, 2007; Hall and Horowitz, 1996), jackknife (Hahn and Newey, 2004; Kezdi, Hahn, and Solon, 2001), analytical methods (Hahn and Kuersteiner, 2002; Hahn and Newey, 2004) and II (Gouriéroux, Phillips and Yu, 2010; Gouriéroux, Renault and Touzi, 2000). Alternatively, one can adjust the estimator to obtain median-unbiased estimators; see e.g. Andrews (1993). With Z_n chosen as the initial estimator, one can think of these methods as a type of GMM procedure where the sample statistic is matched against its model implied version, e.g. its finite-sample mean or median to obtain a new, improved estimator. This is in particular the case with the II estimator when the auxiliary model is chosen as the actual model.

We here propose a method that offer finite-sample improvements over the aforementioned estimation methods. As with all the above GMM-type estimators[1], we take as starting point some statistic Z_n, which, for example, could be an initial estimator of θ_0, a set of sample moments, or an auxiliary model statistic as used in II. However, rather than minimizing some L_2-distance, we propose to (re-)estimate the parameters of interest by

---

[1] We use the term "GMM-type estimators" to refer to GMM, MSM, II or EMM estimators based upon a statistic Z_n, as described in the text.

maximizing the likelihood implied by Z_n. This leads to a maximum-likelihood type estimator which we call the maximum indirect likelihood estimator (MIL), since we operate though the statistic rather than on the sample directly. As a computationally attractive alternative to the MIL, we also propose a Bayesian version of our estimator which is termed a Bayesian indirect likelihood (BIL) estimator. These IL estimators offer finite-sample improvements over the corresponding GMM-type estimators based on the same statistic as we will argue in the following.

We derive the asymptotic distributions of the IL estimators and find that they are first-order equivalent to the GMM estimator that is based on the same auxiliary statistic and uses an optimal weighting matrix. However, the former will in general enjoy better small sample performance compared to the latter for two reasons: First, while GMM estimators only utilize the first and second moment of the statistic, IL estimators are based on a full description of its finite sample distributional characteristics. As such, we expect them to be superior to the GMM estimator in terms of higher-order optimality criteria such as the "large deviations" principle (Bahadur, Zabell and Gupta, 1980), higher-order efficiency (Pfanzagl and Wefelmeyer, 1978), and large deviation probabilities of type II errors (Zeitouni and Gutman, 1991).

Second, the first-order equivalence results rely on the GMM estimator being computed using the optimal weighting matrix. Since this in general is unknown, it has to be estimated in order for the efficient GMM estimator to be feasible. This is particularly difficult in time series models where HAC-type estimators have to be employed. In contrast, for our estimators there is no need to estimate an optimal weighting matrix since the likelihood function already embodies the information inherent in the optimal weight matrix. This eliminates an important source of imprecision that can adversely affect the small sample performance of over identified GMM-type estimators (Altonji and Segal, 1996; Doran and Schmidt, 2006; Hansen, Heaton and Yaron, 1996).

To justify the above claims of higher-order optimality of the MIL over the corresponding GMM estimator, we provide a higher-order asymptotic analysis of both estimators. In particular, we demonstrate that while the competing estimators have same leading variance components, and so are first-order equivalent, the MIL estimator is third-order efficient in the sense that it has a smaller higher-order variance relative to the GMM estimators.

The implementation of the indirect likelihood estimators depends on the likelihood function of the statistic being available on closed form, which will normally not be the case. However, if the model is fully specified, our ability to learn about the likelihood of the statistic is limited only by willingness to do simulations. In particular, we formulate 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 (2009), Fermanian and Salanié (2004), and Kristensen and Shin (2008). The simulated versions are shown to be asymptotically first-order equivalent to the infeasible MIL and BIL estimators as the number of simulations increases.

The above mentioned theoretical arguments for improved finite-sample performance of our indirect likelihood estimators over GMM-type estimators are supported by Monte

Carlo results. We investigate the performance of the proposed estimators using a wide range of models, including time series, dynamic and nonlinear panel data, structural auction, and dynamic stochastic general equilibrium models. In terms of root mean squared error and bias, we find that the simulated version of the BIL estimator exhibits performance that is almost always as good, and in most cases better, than the corresponding GMM-type estimators. In particular, BIL is found to inherit the automated bias-correction feature of the standard Indirect Inference (II henceforth) estimators discussed above.

When this paper was nearly completed, we became aware of so-called Approximate Bayesian Computation (ABC) or likelihood-free Bayesian inference (see, e.g., Tavaré *et al.,* 1997; Marjoram *et al.,* 2003; Sisson, Fan and Tanaka, 2007) which are used in the biological sciences, including genetics, epidemiology and population biology. One form of ABC (Beaumont, Zhang and Balding, 2002) directly implements what we call the simulated BIL (SBIL) estimator. While the ABC literature is quite mature from an empirical point of view, no theoretical results are available for ABC estimators and their simulated versions, and so this paper offers a number of contributions in this direction. Moreover, the ABC literature only contains rather limited results on the BIL's finite-sample performance; we provide extensive Monte Carlo examples investigating this. As such, this paper provides an asymptotic theory and finite-sample analysis that has been missing for this literature.

The remains of the paper is organized as follows: Section 2 presents the indirect likelihood estimators, and Section 3 discusses their implementation. First- and higher-order theory of the estimators are developed in Sections 4 and 5 respectively. Section 6 contains the simulation studies, while Section 7 concludes. All proofs have been relegated to the Appendix.

2. INDIRECT LIKELIHOOD INFERENCE

We consider the setting described in the introduction, where we wish to learn about a parameter $\theta \in \Theta \subset \mathbb{R}^k$ describing a model. Given a sample $Y_n = (y_1, ..., y_n)$ from the model, we choose to make inference on θ through a d-dimensional statistic of the sample, $Z_n = Z_n(Y_n) \in \mathbb{R}^d$. We can think of Y_n as a (random) mapping taking a parameter value into the corresponding observed sample, $Y_n = Y_n(\theta)$. This in turn implies that the statistic also implicitly is a function of θ through the data, and write

$$Z_n(\theta) \equiv Z_n(Y_n(\theta)).$$

In particular, the observed statistic is this random mapping evaluated at the true parameter value which we denote θ_0, $Z_n = Z_n(\theta_0)$. Let $f_n(Z_n|\theta)$ be the likelihood of the statistic for a given value of the parameter. Suppose for now that the likelihood of the statistic is known on closed form.[2] We then propose to estimate the parameters by maximizing the indirect likelihood defined through Z_n:

(1)
$$\hat{\theta}_{MIL} = \arg \sup_{\theta \in \Theta} \log f_n(Z_n|\theta).$$

---

[2]In general, this will not be the case; in the next section we therefore develop a simulated version of it.

The proposed estimator is indirect, because the sample data is filtered through a statistic, and we refer to the estimator as a maximum-indirect likelihood (MIL) estimator.

Compared to the actual MLE based on the full sample, the MIL estimator will in general suffer from an information loss and will only obtain full maximum-likelihood efficiency if the statistic is sufficient in the sense that it spans the score of the full sample log-likelihood. On the other hand, the computation of the indirect likelihood is a lower-dimensional problem compared to the full likelihood ($\dim(Z_n) < \dim(Y_n)$). Moreover, even when the full MLE is computationally feasible, the IL estimator can be used to adjust for finite-sample biases as argued below. Finally, we note that the IL estimator in general will be more robust compared to the full MLE in that it can handle misspecified models and remains consistent as long as Z_n identifies the parameter of interest. These are to some extent shared by GMM estimators based on the same statistic. However, in finite samples the two estimators will perform differently, and the MIL will in general exhibit higher-order improvements relative to the GMM estimator. Two leading examples illustrating this general phenomenon are the following:

In the first example, suppose that we have available some initial estimator, say $\hat{\theta}$. Under suitable regularity conditions, this estimator will be asymptotically normally distributed centered around the true parameter value θ_0. However, in finite samples the estimator will in general not be normally distributed and not be centered around θ_0. It therefore appears sensible to try to learn about the estimator's finite-sample distribution, and utilize this information to obtain a better estimate. By choosing our statistic as $Z_n = \hat{\theta}$, the MIL estimator is an updated version of the initial estimator that takes into account the finite-sample characteristics of $\hat{\theta}$. In particular, we expect that the MIL estimator automatically adjusts for potential biases in the initial estimator. As such it is similar to the II bias adjustment mechanism reported in Gouriéroux, Renault and Touzi (2000) and Gouriéroux, Phillips and Yu (2010). However, since MIL estimator at the same time takes into account features of the distribution of $\hat{\theta}$ beyond its first moment, it should be expected that it will in general dominate the II estimator.

As a second example, suppose Z_n has been chosen as a set of sample moments; this is for example the case with simulated method of moments. These are mean-unbiased estimators of the corresponding population means and so there is no need for bias adjustment. As such it would seem that a (two-step) GMM estimator based on Z_n would suffice. However, the statistic may in finite samples still be non-Normally distributed and taking into account these features will improve the estimator. Furthermore, in the over identified case where $d > k$, the efficient GMM requires either knowledge or a preliminary estimator of the efficient weighting matrix. In contrast, the MIL automatically incorporates information about the efficient weight and as such is similar to the (generalized) empirical likelihood (GEL) estimator in that it utilizes the full distributional characteristics of the chosen statistic in the estimation of the parameters. As a consequence, the MIL estimator will share the higher-order optimality properties of the GEL (see Newey and Smith, 2004) and dominate the corresponding GMM estimator.

In certain situations, the optimization problem defining the MIL estimator may be difficult to solve numerically. The likelihood function $\theta \mapsto f_n(Z_n|\theta)$ may be non convex,

have multiple local maxima, flat spots, or discontinuities, in which case the global maximizer, $\hat{\theta}_{MIL}$, can be difficult to compute in practice. These features may be even more pronounced when the estimation is based on a simulated version of the likelihood function. This is particularly an issue when the parameter space Θ is "large" since the search has to be done over a large-dimensional space. To circumvent these potential problems in the computation of $\hat{\theta}_{MIL}$, we introduce a Bayesian version of it as a computationally attractive alternative, since it does not require numerical optimization. In the simulation studies, we focus on the posterior mean of θ given Z_n defined as

$$(2) \qquad \hat{\theta}_{BIL} = \int_{\Theta} \theta f_n\left(\theta | Z_n\right) d\theta,$$

where $f_n\left(\theta | Z_n\right)$ is the posterior distribution given by

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

for some density $\pi(\theta)$ on the parameter space Θ. We refer to this particular estimator as the Bayesian indirect likelihood (BIL) estimator. More generally, θ_0 could be estimated by:

$$(3) \qquad \hat{\theta}_{BIL} = \arg \inf_{\zeta \in \Theta} \int_{\Theta} \rho\left(\sqrt{n}(\theta - \zeta)\right) f_n\left(\theta | Z_n\right) d\theta,$$

for some penalty or loss function $\rho\left(u\right)$. This includes the posterior mean which is obtained by specifying a quadratic loss $\rho\left(u\right)=|u|^2$, while the τth quantile of the posterior follows from choosing the penalty function as the so-called "check" function, $\rho\left(u\right) = \sum_{i=1}^{k}(\tau_i - \mathbf{1}\{u_i \leq 0\})$, where $\mathbf{1}\{\bullet\}$ denotes the indicator function. The posterior quantiles can be used to construct asymptotically valid confidence intervals as shown in the next section.

It should be stressed that we do not give the BIL estimator a Bayesian interpretation and merely see it as a computational device to circumvent the numerical issues related to the maximization problem that has to be solved in order to compute $\hat{\theta}_{MIL}$. In particular, we do not interpret $\pi(\theta)$ as a prior density in the Bayesian sense, in that it does not necessarily reflect beliefs about the parameter. It is simply used to give weights to different parts of the parameter space, and in our examples, we alway use a uniform density. As such, $\hat{\theta}_{BIL}$ is close in spirit to the class of Laplace type estimators (LTE's) introduced in Chernozhukov and Hong (2003).

3. COMPUTATION OF FEASIBLE ESTIMATORS

In most situations, it will not be possible to derive the exact finite-sample distribution of the statistic Z_n on closed form. Thus the likelihood $f_n(Z_n | \theta)$ will normally not be available, and one has to resort to numerical approximations instead. In the computation of the BIL estimator, it is in addition required to compute the integral $\int_{\Theta} \rho_n\left(\theta - \zeta\right) f_n(\theta | Z_n) d\theta$. For the latter problem, one could follow the suggestions of Chernozhukov and Hong (2003) and compute the integral using Markov chain Monte Carlo (MCMC) methods.

However, we here opt for an alternative solution which handles the numerical approximation of $f_n(Z_n|\theta)$ and the integral in one step; the proposed method which we describe below is easy to implement and in general quite robust.

First, for the implementation of the MIL, we have to be able to compute $f_n(Z_n|\theta)$ at any given trial value θ. Since the model is simulable and the mapping $Z_n(\theta) \equiv Z_n(Y_n(\theta))$ is known (as chosen by the econometrician), we propose to estimate the density using kernel density methods: Draw S independent samples, $Y_n^s(\theta)$ for $s = 1, ..., S$, from the model evaluated at the trial value θ, compute the associated statistic, $Z_n^s(\theta) \equiv Z_n(Y_n^s(\theta))$, $s = 1, ..., S$, and then estimate the density by kernel methods (see e.g. Li and Racine, 2007, Ch. 1 for an introduction):

$$(4) \qquad \hat{f}_{n,S}(Z_n|\theta) = \sum_{s=1}^{S} K_h\left(Z_n^s(\theta) - Z_n\right),$$

where $K_h(z) = K(z/h)/h$, $K(z)$ is a kernel function and $h > 0$ is a bandwidth. One then embeds the approximated density inside (1), and uses an optimization algorithm to obtain an estimator. This yields a simulated MIL (SMIL) estimator:

$$(5) \qquad \hat{\theta}_{SMIL} = \arg\sup_{\theta \in \Theta} \log \hat{f}_{n,S}(Z_n|\theta).$$

The simulated version is akin to the nonparametric simulated maximum-likelihood estimator (NPSMLE) of Fermanian and Salanié (2004) and Kristensen and Shin (2008). The above kernel density estimator implicitly assumes that $Z_n(\theta)$ has a continuous distribution. However, we show that even if this is not the case, the simulated version will still asymptotically behave as the MIL estimator.

For the computation of the BIL estimator, we not only need to evaluate the likelihood but also the integral over the quasi-posterior density. Chernozhukov and Hong (2003) propose to handle the latter computational problem through MCMC, but this can be quite a delicate method which in some cases has unstable properties (see Kormiltsina and Nekipelov, 2009). Instead, we opt to also combine simulations and nonparametric techniques in the implementation of the BIL estimator. Suppose, to illustrate, that the penalty function is $\rho(u) = |u|^2$. In this case, the Laplace-type estimator is the mean of the posterior density,

$$\hat{\theta}_{BIL} = \int_\Theta \theta f_n(\theta|Z_n) d\theta = E[\theta|Z_n].$$

Our idea is then to compute $\hat{\theta}_{BIL} = E[\theta|Z_n]$ by combining nonparametric regression methods and simulations as follows: 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),

$$(6) \qquad \hat{\theta}_{SBIL} = \frac{\sum_{s=1}^{S} \theta^s K_h\left(Z_n^s - Z_n\right)}{\sum_{s=1}^{S} K_h\left(Z_n^s - Z_n\right)},$$

while another one is the *k*-nearest neighbor (KNN) estimator (see Li and Racine, 2007, Ch. 14), where 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. As such the KNN estimator can be thought of as a kernel regression estimator with an adaptive bandwidth.

A member in the general class of BIL estimators given in eq. (3) can be expressed as minimizing a conditional moment,

$$\hat{\theta}_{BIL} = \arg \inf_{\zeta \in \Theta} E \left[\rho \left(\sqrt{n}(\theta - \zeta) \right) | Z_n \right],$$

which can be approximated by replacing the exact moment by a simulated nonparametric version,

$$\hat{\theta}_{SBIL} = \arg \inf_{\zeta \in \Theta} \hat{E}_S \left[\rho \left(\sqrt{n}(\theta - \zeta) \right) | Z_n \right],$$

where $\hat{E}_S \left[\rho \left(\sqrt{n}(\theta - \zeta) \right) | Z_n \right]$, for example, can be computed by kernel regression,

$$(7) \qquad \hat{E}_S \left[\rho \left(\sqrt{n}(\theta - \zeta) \right) | Z_n \right] = \frac{\sum_{s=1}^{S} \rho \left(\sqrt{n}(\theta - \zeta) \right) K_h \left(Z_n^s - Z_n \right)}{\sum_{s=1}^{S} K_h \left(Z_n^s - Z_n \right)},$$

or nearest neighbor estimation where again $h = d_k(Z_n)$. When the penalty function is chosen as the "check"-function, this leads to simulated versions of the posterior quantiles, which are used to compute confidence intervals. In this case, the kernel smoothed version becomes the kernel quantile regression estimator (Li and Racine, 2007, Sec. 6.4).

For both the SMIL and SBIL, there are two sources of error in comparison with the exact MIL and BIL estimators (which only suffer from the sampling error in Z_n). First, randomness is added due to the use of simulations, and there is also a bias component due to the use of nonparametric estimators. We treat the nonparametric fitting step as a computational tool used to find the value of the estimator, in the same way that Chernozhukov and Hong (2003) treat MCMC as a means of computing LTEs. As the number of simulated draws S becomes large, nonparametric density and regression estimators are consistent. Thus, both the randomness due to use of simulations and the bias due to use of nonparametric methods can be controlled for by choosing S sufficiently large. We analyze the impact of simulations and kernel smoothing in Section 6.

One may wish to explore different pseudo-priors. If a large body of simulations have been generated using the pseudo-prior $\pi(\theta)$, then one can obtain results for a different pseudo-prior without doing additional simulations by using importance sampling. The SBIL estimator based on $\pi(\theta)$ presented in equation 6 can be written as $\hat{\theta}_{SBIL} = \sum_{s=1}^{S} \theta^s w_S(Z_n^s, Z_n)$, where $w_S(Z_n^s, Z_n)$ has an obvious definition. Given simulations $\{(\theta^s, Z_n^s)\}_{s=1}^{S}$ based on $\pi(\theta)$, the SBIL estimator corresponding to the new pseudo-prior, say $\pi^*(\theta)$, can be computed by

$$\hat{\theta}_{SBIL} = \sum_{s=1}^{S} \theta^s w_S(Z_n^s, Z_n) \frac{\pi^*(\theta^s)}{\pi(\theta^s)}.$$

This may be useful when S is very large or when it is costly to compute the auxiliary statistic, as in the case of the DSGE models presented later in this paper.

In the ABC literature or likelihood-free literature, discussed in the introduction, methods of computing estimators using likelihood-free Markov chain Monte Carlo and sequential Monte Carlo have been studied in some detail (Marjoram *et al.,* 2003; Sisson, Fan and Tanaka, 2007; Beaumont *et al.* 2009). One could also employ so-called importance sampling to reduce variances due to simulations: For any conditional density $g_n(\theta|z)$ with support Θ, we can rewrite $\hat{\theta}_{BIL}$ as $\hat{\theta}_{BIL} = \int \theta \{f_n(\theta|Z_n)/g_n(\theta|Z_n)\} g_n(\theta|Z_n) d\theta$, and so a generalized version of our proposed simulated version would be

$$\hat{\theta}_{BIL} = \frac{1}{S} \sum_{s=1}^{S} \theta^s \frac{\hat{f}_{n,S}(\theta^s|Z_n)}{g_n(\theta^s|Z_n)} = \frac{\sum_{s=1}^{S} \theta^s \pi(\theta^s)/g_n(\theta^s|Z_n) K_h(Z_n^s - Z_n)}{\sum_{s=1}^{S} K_h(Z_n^s - Z_n)},$$

where $\theta^s \sim$ i.i.d. $g_n(\theta^s|Z_n)$. The optimal choice of $g_n(\theta|z)$ in terms of variance reduction is

$$g_n(\theta|z) = \frac{\mathbf{1}\{\theta \in \Theta\} |\theta| f_n(\theta|z)}{\int_\Theta |\theta| f_n(\theta|z) d\theta}.$$

Unfortunately, it is not feasible to draw from this choice since $\int_\Theta |\theta| f_n(\theta|z) d\theta$ is unknown, but approximate methods exist; see, for example, Zhang (1996). These methods should in principle be computationally more efficient compared to the basic sampling method in equation (6) to obtain a given level of precision. In our simulation study we focus on the basic sampler, and leave the implementation of importance samplers for future research. Application of these methods could provide savings in computational time when it is costly to sample from the model, but on the other hand require more careful implementation. Of the examples considered in this paper, only the DSGE models (below) present serious computational burden.

4. First-order Asymptotics

As a first step towards a complete asymptotic analysis of the MIL and BIL estimators, we here derive their first-order asymptotic distribution. The asymptotic analysis of the MIL estimator proceeds along the standard steps for parametric extremum estimators, while the BIL estimator on the other hand requires a bit more care. Fortunately, since the BIL estimator can be regarded as a specific LTE, we can employ the general results of Chernozhukov and Hong (2003) to establish $\sqrt{n}$-consistency and asymptotic normality of our Bayesian estimator, as well as the equivalence with the MIL estimator when the penalty function ρ is symmetric.

We impose the following conditions on the parameter space and the weighting function

Assumption 1. *Assume that: (i) the parameter space $\Theta \subset \mathbb{R}^k$ is compact with θ_0 being an interior point; (ii) the weighting function $\pi(\theta)$ is a continuous, uniformly positive density; and (iii) the penalty function is convex and satisfies $\rho(u) = 0 \Leftrightarrow u = 0$, $\rho(u) \leq 1 + |u|^p$ for some $p \geq 1$, and $\phi(x) = \int \rho(u - x) e^{u'au} du$ is uniquely minimized at some x^* for any $a > 0$.*

This set of assumptions is completely standard, and are identical to the conditions found in Chernozhukov and Hong (2003). It should be noted that (ii)-(iii) are only needed to develop theory for the BIL estimator, and the asymptotics of the MIL estimator only require (i).

Next, we restrict our attention to statistics that are asymptotically normally distributed around a limit $Z(\theta_0)$:

Assumption 2. *The sample statistic $Z_n = Z_n(\theta_0)$ satisfies $\sqrt{n}(Z_n - Z(\theta_0)) \to^d N(0, \Omega(\theta_0))$ for some vector $Z(\theta) \in \mathbb{R}^d$ and covariance matrix $\Omega(\theta) \in \mathbb{R}^{d \times d}$.*

This assumption covers most known statistics in regular, stationary (in particular, cross-sectional) models. In particular, if the statistic is chosen as a set of sample moments, we can appeal to a Central Limit Theorem (CLT) which will hold for stationary processes under suitable mixing and moment conditions. If the statistic is a preliminary estimator, then the estimator can (in large samples) be represented as a set of moments and again a suitable version of the CLT can be employed to verify the assumption.

To ensure that the parameter θ is identified through the statistic in the population, we assume the following regarding its asymptotic mean and variance:

Assumption 3. *The 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) $J(\theta_0) := \dot{Z}(\theta_0)' \Omega^{-1}(\theta_0) \dot{Z}(\theta_0)$ has full rank where $\dot{Z}(\theta) = \partial Z(\theta) / (\partial \theta') \in \mathbb{R}^{d \times k}$.*

This assumption is fairly standard and are 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, $J^{-1}(\theta_0)$ is the asymptotic variance of the IL estimators with $\dot{Z}(\theta_0)$ capturing the information content of the chosen statistic and $\Omega(\theta_0)$ the finite sample variation of it.

When the statistic is chosen as an extremum estimator of an auxiliary model, $Z(\theta)$ plays the role of the so-called "binding" function between the auxiliary model and the parameter of interest, and Assumption 3 is the usual requirement that this is one-to-one.

In the case where the statistic has been chosen as an initial estimator, Assumption 3 holds with $Z(\theta) = \theta + B(\theta)$ where $B(\theta)$ is the asymptotic bias. If the estimator is consistent ($B(\theta) = 0$), Assumptions 3(i)-(ii) are automatically satisfied if the model is specified such that no stochastic singularities are present. Moreover, the asymptotic variance of the IL estimators is in this case equal to the initial estimator's, $J^{-1}(\theta_0) := \Omega(\theta_0)$. Thus, the IL estimators obtains full maximum-likelihood efficiency if the preliminary estimator is (asymptotically) equivalent to the MLE based on the full sample. If the preliminary estimator is asymptotically biased, we require that $B(\theta)$ is one-to-one.

Another way of formulating Assumption 2 is that the cumulative distribution function (cdf) of $Z_n(\theta)$, $F_n(z|\theta) := P(Z_n(\theta) \leq z)$, satisfies

$$(8) \qquad F_n(z|\theta) = P(T_n(\theta) \leq T_n(z|\theta)) = \Phi(T_n(z|\theta)) + o(1),$$

where Φ denotes the cdf of a standard normal distribution, and $T_n(\theta)$ is the normalized statistic,

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

One would therefore expect the MIL estimator to be asymptotically (first-order) equivalent to the MLE based on the Gaussian likelihood given by

$$
\phi_n^* (z|\theta) := \frac{\partial \Phi \left(T_n \left(z|\theta \right) \right)}{\partial z} = \sqrt{n / |\Omega (\theta)|} \phi \left(T_n \left(z|\theta \right) \right). \tag{9}
$$

It could be tempting to take derivatives on both sides of eq. (8) and conclude this is case. Unfortunately, weak convergence does not necessarily imply convergence of corresponding likelihood, so this argument is not necessarily correct. First of all, if $Z_n (\theta)$ has a discrete distribution, its likelihood cannot be found as the derivative of $F_n (z|\theta)$. And even if the likelihood can be described through a density such that $f_n (z|\theta) = \partial F_n (z|\theta) / (\partial z)$, it may not necessarily converge. To resolve these issues, we directly assume that the distribution of the normalized statistic $T_n (\theta)$ satisfies an Edgeworth expansion based on a limiting normal distribution; see Hall (1992) for an introduction to these. Let $f_{T_n} (t|\theta)$ denote the likelihood of $T_n (\theta)$. We then assume that:

Assumption 4. *The normalized statistic $T_n (\theta)$ satisfies an Edgeworth expansion of order $r \geq 0$ uniformly in θ:*

$$
\sup_{t \in \mathbb{R}^d} \left| f_{T_n} (t|\theta) - f_{T_n}^* (t|\theta) \right| = o \left(n^{-r/2} \right),
$$

where

$$
f_{T_n}^* (t|\theta) = \phi (t) \left[1 + \sum_{i=1}^{r} n^{-i/2} \pi_i (t|\theta) \right], \tag{10}
$$

and $t \to \pi_i (t|\theta)$ is a polynomial of order $3i$ with coefficients that are smooth in θ, $i = 1, ..., r$.

Assumption 4 is quite high-level, but is satisfied under great generality since most regular statistics satisfy an Edgeworth expansion. Suppose first that the sample is i.i.d. and Z_n is a sample average, $Z_n = \sum_{i=1}^{n} g (y_i) / n$, with $g (y_i)$ having a continuous distribution. Then the above Edgeworth expansion holds under weak regularity conditions; see Hall (1992, Section 2.8). If the statistic 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, $T_n (\theta)$ defined above, still satisfies eq. (10), see Bhattacharya and Ghosh (1978). Edgeworth expansions are also available for estimators with dependent data under suitable conditions on the dependence structure, see e.g. Fuh (2006), Hall and Horowitz (1996), Inoue and Shintani (2006). Finally, amongst others, Phillips (1977) and Skovgaard (1981,1986) give general conditions under which transformations of Edgeworth expandable statistics themselves have Edgeworth expansions. Thus, Assumption 4 holds for a wide range of relevant statistics when data follows a continuous distribution.

If the underlying observations are discretely distributed, the likelihood of Z_n will in general still be well-defined and an Edgeworth expansion of the cumulative distribution function of Z_n will still hold (see, for example, Bhattacharya and Rao, 1976), but the theory gets less tractable. We conjecture that Assumption 4 will still hold in this case.

For the first-order analysis, the assumption of an Edgeworth expansion can be replaced by alternative, more primitive conditions. In the next section, where we analyze the properties of the simulated versions of the MIL and BIL estimators, we demonstrate that Assumption 4 can be replaced by a tightness condition which holds under great generality for both continuous and discrete random variables.

Finally, we note that Assumption 4 implies Assumption 2 if the model is correctly specified. However, since we want to allow for (moderate) misspecifications where the model is correct only so far that the limit of $Z_n(\theta)$ identifies θ, we maintain both assumptions. This allows us to show that the MIL and BIL estimators share the well-known robustness feature of GMM estimators: They remain consistent as long as the statistic converges towards a limit that identifies the parameters.

Assumption 4 allows us to formalize the intuition given above that $f_n(z|\theta)$ should be well-approximated by the corresponding Gaussian likelihood. To see this, let $f_n^*(z|\theta)$ denote the corresponding rth order approximation of the likelihood of Z_n,

$$(11) \qquad f_n^*(z|\theta) = \sqrt{\frac{n}{|\Omega(\theta)|}} f_{T_n}^*(T_n(z|\theta)|\theta) = \phi_n^*(z|\theta) \left[1 + \sum_{i=1}^{r} n^{-i/2} \pi_i(T_n(z|\theta)|\theta) \right].$$

Assumption 4 then implies:

$$\sup_{z \in \mathbb{R}^d} |f_n(z|\theta) - f_n^*(z|\theta)| = \sqrt{\frac{n}{|\Omega(\theta)|}} \sup_{z \in \mathbb{R}^d} \left| f_{T_n}(T_n(z|\theta)|\theta) - f_{T_n}^*(T_n(z|\theta)|\theta) \right| = o_P\left(n^{-(r-1)/2}\right).$$

In particular, for n large enough, $\frac{1}{2}\phi_n^*(z|\theta) \le f_n(z|\theta) \le 2\phi_n^*(z|\theta)$. Using this bound in conjunction with a first order Taylor expansion, it holds for any $B > 0$,

$$
\begin{aligned}
\sup_{|z| \le B, \theta \in \Theta} |\log f_n(z|\theta) - \log f_n^*(z|\theta)| &\le& \sup_{|z| \le B, \theta \in \Theta} \frac{1}{2 f_n(z|\theta) + o_P(1)} |f_n(z|\theta) - f_n^*(z|\theta)| \\
&\le& \frac{C}{\sqrt{n}} exp\left[nB^2\right] \sup_{|z| \le B, \theta \in \Theta} |f_n(z|\theta) - f_n^*(z|\theta)|,
\end{aligned}
$$

for some constant $C < \infty$. Since $Z_n = Z(\theta_0) + O_P(1/\sqrt{n})$, we can choose the bound $B = B_n = B_0 n^{-1/2}\log(n)$ for some $B_0 > 0$ and obtain in total that

$$\sup_{\theta \in \Theta} |\log f_n(Z_n|\theta) - \log f_n^*(Z_n|\theta)| = o_P\left(\log(n) n^{-(r-1)/2}\right).$$

We state this as a lemma:

Lemma 1. *Under Assumptions 1-4, the (pseudo-)log-likelihood ratio,*

$$LR_n(\theta) := \frac{1}{n} \left\{ \log f_n(Z_n|\theta) - \log f_n^*(Z_n|\theta) \right\},$$

satisfies $|\log LR_n(\theta)| = o_P\left(\log(n) n^{-(r+1)/2}\right)$ *uniformly over* θ.

We are now able to derive the first-order properties of the MIL and BIL estimators through the following two steps: First, by Lemma 1, the IL estimators are first-order equivalent to the maximizer of the Gaussian approximation. Second, the maximizer of the Gaussian approximation is $\sqrt{n}$-asymptotically normally distributed. This yields the following result:

Proposition 1. *Under Assumptions 1-4 with the loss function ρ being symmetric and $r \geq 1$,*

$$\sqrt{n}(\hat{\theta}_{MIL} - \theta_0) \to^d N\left(0, J^{-1}(\theta_0)\right) \text{ and } \sqrt{n}(\hat{\theta}_{BIL} - \theta_0) \to^d N\left(0, J^{-1}(\theta_0)\right),$$

where $J(\theta_0)$ is defined in Assumption 3.

As expected, the two IL estimators are first-order asymptotically equivalent as is standard for frequentist and Bayesian versions of the same estimator.

One could say that the IL estimators exhibit "limited" maximum-likelihood efficiency since their asymptotic variance is optimal given the statistic. However, this does not necessarily mean that the MIL and BIL estimators are equivalent to the MLE based on the full sample. For this to happen, the chosen statistic has to be sufficient in the sense that it has to span the score of the full likelihood, $f(Y_n|\theta)$. On the other hand, the MIL and BIL estimators have a number of attractive features over the full MLE as discussed earlier.

The above result allows one to draw inference regarding the parameter and confidence intervals can for example be computed in the standard way given an estimator of the asymptotic variance, $J^{-1}(\theta_0)$. A standard estimator method would be to utilize the sandwich form of $J(\theta_0)$ 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 the standard estimator of the information of a MLE,

$$\hat{J} = \frac{1}{n} \frac{\partial^2 \log f_n(Z_n|\theta)}{\partial \theta \partial \theta'}\bigg|_{\theta=\hat{\theta}},$$

where $\hat{\theta}$ is a consistent estimator of θ_0 such as either the MIL or BIL. This can be obtained by taking either numerical derivatives of the log-likelihood or alternatively one can try to derive an explicit form of the first and second order derivative of $Z_n(\theta)$ in which case the second derivative of the kernel estimator w.r.t. θ can be used to estimate $J(\theta_0)$. Finally, consistent confidence bands can also be computed using the posterior quantiles. An application of Theorem 3 of Chernozhukov and Hong (2003) shows this:

Proposition 2. *Under Assumptions 1-4, quantiles of the posterior distribution can be used to construct a confidence interval that has proper asymptotic coverage.*

The above asymptotic first-order analysis of the IL estimators relies on the assumption that the likelihood function converges towards its Gaussian approximation, $\phi_n^*(z|\theta)$. One would therefore expect that estimators defined directly in terms of the Gaussian approximation will be first-order equivalent to the IL estimators. Maximizing $\log \phi_n^*(z|\theta)$ is equivalent to minimizing $\frac{1}{2}(Z_n - Z(\theta))' \Omega^{-1}(\theta)(Z_n - Z(\theta))$ which we recognize as the objective function of the continuous updating estimator (CUE) as proposed in Hansen, Heaton and Yaron (1996) when the form of the optimal weighting matrix $\Omega^{-1}(\theta)$ is known. We now give a formal proof of that the IL estimators are asymptotically first-order equivalent to this (ideal) CUE. However, since we also wish to include two-step estimators and standard II estimators in the class of GMM estimators, we introduce a

more general minimum-distance objective function:

$$(12) \qquad D_n(\theta) = \frac{1}{2} \left(Z_n - \bar{Z}_n(\theta) \right)' W_n \left(Z_n - \bar{Z}_n(\theta) \right).$$

where $\bar{Z}_n(\theta)$ plays the role of the binding function which is allowed to depend on sample size, and W_n is some positive definite weighting matrix. The binding function could, for example, be the finite-sample mean of the statistic, $\bar{Z}_n(\theta) = E_\theta[Z_n]$, but other options are allowed for such as its finite-sample median or the asymptotic limit of the finite sample moment. The corresponding GMM estimator is then given as

$$(13) \qquad \hat{\theta}_{GMM} = \arg\min_{\theta \in \Theta} D_n(\theta).$$

The following proposition states that the CUE and the efficient two-step GMM estimator are asymptotically first-order equivalent to the MIL and BIL estimators:

Proposition 3. *Under Assumptions 1-3, the two-step GMM estimator defined in eqs. (13)-(12), where $\bar{Z}_n(\theta) = Z(\theta) + o\left(1/\sqrt{n}\right)$, with the weight matrix satisfying $W_n = \Omega^{-1}(\theta_0) + o_p(1)$ is first-order equivalent to the MIL and BIL: $\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \to^d N\left(0, J^{-1}(\theta_0)\right)$.*

Given the above result, one may at this stage then ask what the advantages of IL estimators over standard GMM estimators are? This is answered in the next section, where we demonstrate that the IL estimators are higher-order efficient compared to the above class of GMM estimators.

5. Higher-Order Asymptotics

In this section, we develop higher-order theory for the IL estimator and their GMM counterparts introduced in the previous section. The purpose with this higher-order analysis is two-fold: First, we analyze the bias properties of the IL estimators and compare those with the ones of the GMM type estimators. Second, we show that the MIL estimator is higher-order efficient compared to the competing GMM estimator. To keep the notation at a reasonable level and to avoid overly complicated proofs, we restrict ourselves to the case of a scalar parameter, $\theta \in \mathbb{R}$. Moreover, we focus on the MIL in the following since maximum-likelihood and Bayesian estimators are, in general, higher-order equivalent (see, for example, Gusev, 1975), and so we expect that the analysis of the MIL carries over to the BIL estimators.

In the following we assume that the model-implied function $\bar{Z}_n(\theta)$ used in the definition of the GMM estimator in eq. (13) is chosen as the model-implied finite-sample moment of Z_n, $\bar{Z}_n(\theta) = E_\theta[Z_n]$. Assuming that higher-order moments of Z_n exist, Assumption 4 with $r \geq 1$ implies that

$$\bar{Z}_n(\theta) = Z(\theta) + B(\theta)/n + o(1/n),$$

$$E_\theta\left[\left(Z_n - \bar{Z}_n(\theta)\right)\left(Z_n - \bar{Z}_n(\theta)\right)'\right] = \Omega(\theta)/n + o(1/n),$$

and $E_\theta\left[\left|Z_n - \bar{Z}_n(\theta)\right|^3\right] = O\left(1/n^{3/2}\right)$. We use these properties in the analysis of the bias properties of the estimators. We first obtain an expansion of the MIL, GMM and CU estimators along the lines of Newey and Smith (2004). For each of the three estimators

(two-step GMM, CU and MIL), we show that the following expansion holds:

(14) $$\hat{\theta} = \theta_0 + \psi_n + Q_1(\psi_n, A_n) + Q_2(\psi_n, A_n, B_n) + R_n,$$

where Q_1 is quadratic in its arguments, Q_2 is cubic, and the remainder term $\tilde{R}_n = O_P(1/n^2)$. Ignoring the remainder term and taking expectations on both sides of the above equation, we obtain the following result regarding the bias properties of the estimators:

Proposition 4. *Assume that Assumptions 1-4 hold with $r \geq 3$, $E\left[|Z_n|^3\right] < \infty$ for all $n \geq 1$ and $\Delta_n := W_n - \Omega_n^{-1}(\theta_0) = O_P(1/\sqrt{n})$. Then,*

$$E\left[\hat{\theta}_{CU}\right] - \theta_0 \simeq \frac{1}{6n}J^{-2}(\theta_0)D^2\bar{m},$$

$$E\left[\hat{\theta}_{GMM}\right] - \theta_0 \simeq \frac{1}{n}J^{-2}(\theta_0)\left\{\frac{1}{6}D^2\bar{m} + B_{W,n}\right\},$$

$$E\left[\hat{\theta}_{MIL}\right] - \theta_0 \simeq \frac{1}{n}J^{-2}(\theta_0)\left\{\frac{1}{6}D^2\bar{m} + B_{\pi}\right\},$$

where, with $\bar{\pi}_{1,i}^{(3)}(t|\theta) = \partial\pi_1(t|\theta) / (\partial t \partial t' \partial t_i)\,\Omega_i^{-1/2}(\theta)$,

(15) $$D^2\bar{m} = -3\ddot{Z}(\theta_0)\,\Omega^{-1}(\theta_0)\,\dot{Z}(\theta_0),$$

(16) $$B_{W,n} = -\dot{Z}(\theta_0)'\,E\left[\Delta_n\dot{Z}(\theta_0)(Z_n - Z(\theta))\right]'\Omega^{-1}(\theta_0)\,\dot{Z}(\theta_0),$$

$$B_{\pi} = \sum_i \dot{Z}(\theta_0)'\,\Omega^{-1/2}(\theta_0)\,\bar{\pi}_{3,i}^{(3)}(0|\theta_0)\,\Omega^{-1/2}(\theta_0)\,\dot{Z}(\theta_0).$$

The expressions of the biases for the CU and two-step GMM estimators are on the same form as in Newey and Smith (2004). We see that in terms of bias the CU estimator will in general dominate both the two-step and MIL estimator: The bias component $B_{W,n}$ for the two-step estimator is due to the estimation of the efficient weighting matrix as is also found in Newey and Smith (2004), while B_{π} captures the curvature of the non-Gaussian component of the indirect likelihood function. In particular, if the indirect likelihood function is close to being Gaussian, $\bar{\pi}_{1,i}^{(3)}(0|\theta_0) = 0$ which in turn implies that $B_{\pi} = 0$. In this case, the MIL has the same first-order bias as the CU estimator. Our simulation results shows that in practice the bias of the MIL is in most cases very similar if not smaller than that of the CU estimator for reasonable sample sizes. This finding indicates that indeed the indirect likelihood is sufficiently close to its Gaussian approximation such that the additional bias term B_{π} is negligible. An alternative explanation of this finding is that higher-order bias terms not included in the above analysis cancel out parts of the leading bias term in finite samples.

Consider the case where the statistic is chosen as an initial estimator, say $Z_n = \tilde{\theta}$. For this choice, the GMM estimator and CUE correspond to the Indirect Inference estimator where the auxiliary model is identical to the actual model. As is well-known, the II estimator in this case automatically bias adjusts. Formally, this follows from the above proposition since in this case $Z(\theta) = \theta$ and so $\ddot{Z}(\theta_0) = 0$ such that the CUE has bias of order $O(1/n^2)$. The MIL does not have this property in general since B_{π} remains different from zero for this special case. However, as mentioned earlier, through simulations

we demonstrate that the bias properties of the MIL are very favorable and are as good as the ones of the CU GMM-type estimator.

We now turn our attention to the higher-order efficiency of the GMM and IL estimators. As was shown in the previous section, their first-order asymptotic variances are identical. However, in finite samples, the IL estimators are expected to dominate for a number of reasons: First, the GMM estimators are only first-order equivalent to the IL estimators if $W_n = \Omega^{-1}(\theta_0) + o_p(1)$. If not, the IL estimators are asymptotically more efficient than GMM. Moreover, the first-step estimation error contained in W_n in general has an adverse impact on the performance of the resulting two-step estimator which may perform poorly in small and moderate samples; see e.g. Altonji and Segal (1996), Hansen, Heaton and Yaron (1996) and Newey and Smith (2004). Furthermore, while increasing the dimension of the auxiliary statistic increases the asymptotic efficiency of the GMM estimator, it also increases the dimension of the weight matrix to be estimated and numerical singularities can appear making the inversion difficult. In contrast, our estimators do not require estimation of the optimal weighting matrix, and so increasing the dimension of the auxiliary statistic causes no difficulties with singular matrices.

If on the other hand $\Omega(\theta)$ is known, then we can estimate the parameters using the CU estimator which will remove the additional estimation errors due to the use of W_n; see Donald and Newey (2000) and Newey and Smith (2004). However, in finite samples, the CUE still only utilizes information contained in the first and second moments of Z_n, while the MIL takes into account all distributional characteristics. This difference means that the indirect likelihood estimators in general will have better small sample performance than both two-step efficient GMM and CU based on the same auxiliary statistic.

The formal proof of higher-order efficiency can be done by ranking the GMM-type and MIL estimators in terms of their higher-order MSE. If an estimator $\hat{\theta}$ satisfies the expansion in eq. (14), we obtain (again ignoring R_n)

$$MSE\left(\sqrt{n}(\hat{\theta} - \theta_0)\right) \simeq B_n B_n' + V_n,$$

where $B_n = \sqrt{n}\text{Bias}\left(\hat{\theta}\right)$ and $V_n = n\text{Var}\left(\hat{\theta}\right)$. For each of the three estimators, the variance can be decomposed into

$$V_n = J^{-1} + \Xi/n + o\left(1/n\right),$$

where $J^{-1} = J^{-1}(\theta_0)$ is the leading variance component, while Ξ is the higher-order variance. The expression of Ξ for each of the three estimators (CUE, GMM, MIL) is straightforward to obtain from the expansion, but it is rather complicated. This makes a direct ranking of the estimators in terms of their respective Ξ's difficult.

Instead, we first develop an Edgeworth expansion of the distribution of the MIL estimator. For standard maximum-likelihood estimators where the log-likelihood takes the form of a sample average over i.i.d. observations, Edgeworth expansions have been established; see, for example, Bhattacharya and Ghosh (1978). However, 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, by importing some of the arguments of Bhattacharya and Ghosh (1978), we can still show that $\hat{\theta}_{MIL} \simeq H(W_n(Z_n))$ for some analytic function H and with $W_n(Z_n)$ denoting the first r derivatives of $\log f_n(Z_n|\theta)$ w.r.t.

θ. Since (the normalized version of) Z_n satisfies an Edgeworth expansion, we can then apply the general results of Phillips (1977) on Edgeworth expansions of transformations of random sequences to obtain the desired result:

Proposition 5. *Under Assumptions 1-4 with $E\left[|Z_n|^p\right] < \infty$ for all $n, p \geq 1$, and*

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

the MIL satisfies an rth order Edgeworth expansion:

$$\sup_y \left| P\left(\sqrt{n}\left(\hat{\theta}_{MIL} - \theta_0\right) J(\theta_0) \leq y\right) - \int_{-\infty}^{y} \phi(x)\left[1 + \sum_{i=1}^{r} n^{-i/2}\tilde{\pi}_i(x)\right] dx \right| = o\left(n^{-r/2}\right),$$

where $\tilde{\pi}_i(x)$ is a polynomial of order $3i$, $i = 1, ..., r$.

The assumption that Z_n has moments of all orders is somewhat restrictive and rules out heavy tails. We conjecture that this assumption is not strictly necessary for the above result to holds. In particular, one might be able to show Proposition 5 by using the results of Skovgaard (1981) where weaker moment restrictions are required. This would on the other hand complicate the proof and so for clarity we maintain the assumption of all moments existing. The tail probability condition is satisfied for most regular statistics; see, e.g., Bhattacharya and Ghosh (1978, Theorem 3).

Once we have shown that the distribution of the MIL estimator can be approximated by an Edgeworth expansion, it now follows by standard results for maximum-likelihood estimators (see e.g. Ghosh, 1994 and Bickel, Götze, and van Zwet, 1985), that the bias-adjusted MIL estimator is third-order efficient amongst all estimators relying on the statistic Z_n. In particular, $\Xi_{GMM} \geq \Xi_{MIL}$ and $\Xi_{CUE} \geq \Xi_{MIL}$.

6. PROPERTIES OF SIMULATED VERSIONS

We analyze the impact of the use of simulations and nonparametric estimation in the implementation of the MIL and BIL estimators. For the simulated version of the MIL estimator, we combine the general results of Kristensen (2009) and Kristensen and Shin (2008) to show that it is first-order asymptotically equivalent to the infeasible MIL estimator. The analysis of the simulated version of the BIL estimator can be done directly since we can write it up on closed form.

To utilize existing results on convergence rates of kernel estimators, we make the following assumptions regarding the kernel function used in the computation of the SMIL and SBIL defined in Section 3:

Assumption 5. *The kernel K satisfies: There exist $C, L < \infty$ such that either (i) $K(u) = 0$ for $\|u\| > L$ and $|K(u) - K(u')| \leq C\|u - u'\|$, or (ii) $K(u)$ is differentiable with $sup_u |K'(u)| < \infty$. For some $a > 1$, $|K(u)| \leq C\|u\|^{-a}$ for $\|u\| > L$, and $\int K(z)\,dz = 1$, $\int zK(z)\,dz = 0$, $\int z^2 K(z)\,dz < \infty$.*

The above assumptions imposed on the kernel are quite standard and are for example satisfied by the Gaussian kernel. We first restrict ourselves to the case where the likelihood is a density:

Assumption 6. *The indirect likelihood $f_n(z|\theta)$ is a density with respect to the Lebesgue measure and is twice continuously differentiable in z.*

Under this assumption on the kernel and the likelihood, the following result holds:

Proposition 6. *Assume that Assumptions 1-6 hold. Then the SMIL and SBIL estimators are asymptotically first-order equivalent to the actual ones under the following conditions:*
For the kernel-smoothed versions, $nh^2 \to 0$ and $n\log(S) / (Sh^d) \to 0$.
For the nearest-neighbor versions, $n[k/S]^{2/d} \to 0$, and $n\log(S)/k \to 0$.

The restrictions on S and h are fairly standard and require the number of simulations to grow at a slightly faster rate than the number of observations. In particular, standard bandwidth selectors will satisfy the above rates and so these can be used in the implementation of the simulated versions.

We also note that the simulated versions of our estimators suffer from a curse of dimensionality. This appears explicitly in the conditions on S and h given in Proposition 6 where we require $n\log(S) / (Sh^d) \to 0$. Thus, the larger $d = \dim(Z_n)$ (which must be at least that of θ, and which is larger in most of the applications below), the more simulations are required for the simulations to have a negligible impact on the estimator. This is a well-known issue which is shared by most other simulation-based estimators: The larger the dimension of the space over which we need to integrate, the larger the number of simulations should be chosen to control the simulation error.

The above result requires the likelihood to be a density. We now demonstrate that the SMIL and SBIL estimators enjoy the same asymptotic properties even if this is not the case. In fact, we will not even require that Assumption 4 holds and as such allow for both continuous and discrete observations. To be more specific, we replace Assumptions 4 and 6 with the following one:

Assumption 7. *For some $N \geq 1$: $sup_{n \geq N} E\left[sup_{\theta \in \Theta} |\sqrt{n}(Z_n(\theta) - Z(\theta))|^2\right] < \infty.$*

This uniform integrability assumption is satisfied if, for example, $Z_n(\theta)$ is a sample average with second moment. It imposes no smoothness restrictions on the finite-sample likelihood and holds for both continuous and discrete underlying data. It is used in conjunction with Assumption 2 to ensure that $\sqrt{n}E[|Z_n(\theta) - Z_n^*(\theta)|] \to 0$, where $Z_n^*(\theta) \sim N(Z(\theta), \Omega(\theta)/n)$ is its Normal limit sequence. We use this to show that the kernel smoother based on simulations from the distribution of $Z_n(\theta)$ converges towards the one based on simulations of $Z_n^*(\theta)$. Since $Z_n^*(\theta)$ satisfies Assumption 4 and 6 by construction, this in turn implies that SMIL and SBIL have the desired asymptotic properties:

Proposition 7. *Assume that Assumptions 1-3, 5 and 7 hold, and the kernel K is uniformly Lipschtiz, $|K(u) - K(v)| \leq D|u - v|$. Then the SMIL and SBIL have the same asymptotic properties as those stated in Proposition 1 under the bandwidth conditions stated in Proposition 6 together with $nh^2 \to \infty$ (kernel smoother) and $n/k^2 \to \infty$ (nearest neighbor).*

The intuition behind the above result is the following: If the distribution of $Z_n(\theta)$ cannot be described by a density, one can think of the kernel smoothing inherent in both the SMIL and SBIL as a type of regularization that generates a smooth objective function

which can be used instead of the more irregularly behaved true likelihood. As such the SMIL and SBIL estimators are similar in nature to the smoothed maximum score estimator proposed in Horowitz (1992) where a non-smooth estimator is regularized through smoothing.

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 is independent of $\hat{\theta}_{BIL}$ and satisfies either (in the case of kernel-smoothers),

$$\sqrt{Sh^d} E_S(z) \to^d N\left(0, \|K\|^2 \frac{\sigma_n^2(z)}{f_n(z)}\right),$$

or (in the case of nearest-neighbor estimators),

$$\sqrt{k} E_S(z) \to^d N\left(0, \|K\|^2 \sigma_n^2(z)\right),$$

where $d = dim(Z_n)$, $\|K\|^2 = \int K^2(z)\,dz$, and $\sigma_n^2(z) = Var[\theta|Z_n = z]$. Thus, the variance estimator of the kernel-smoothed version of SBIL could be adjusted by adding $\|K\|^2 \frac{\sigma_n^2(z)}{f_n(z)} / (Sh^d)$ to $J^{-1}(\theta_0)$, and similarly for the nearest-neighbor version. A similar adjustment can be developed for the MIL estimator by using the arguments of Kristensen and Salanié (2010).

7. MONTE CARLO RESULTS

In this section we explore the performance of the SMIL and SBIL estimators, comparing them to other estimators, using a variety of econometric models including simple time series models, a dynamic and nonlinear panel data models, a structural econometric model of an auction and two dynamic stochastic general equilibrium (DSGE) models. We focus on several issues. First, the SBIL estimator is considerably more convenient to use than is the SMIL estimator, from a computational point of view, so we would like to know if the two estimators perform similarly before focusing our attention on the SBIL estimator. Second, Proposition 5 tells us that the exact MIL is higher-order more efficient than the GMM estimator that uses the optimal weight matrix. This leads us to hope that the SMIL and SBIL estimators have better small sample performance than GMM-type competitors. A factor that could undermine these potential gains is the need to use simulations and nonparametric fitting to implement the feasible versions (the feasible SMIL and SBIL versus the infeasible MIL and BIL). This section throws light on the actual performance of the feasible versions. A third purpose of this section is simply to give examples of how the SMIL and SBIL estimators may be implemented in practice. Examples of practical issues to deal with are the choice of the auxiliary statistic, and the specification of the parameter space in the case of the SBIL estimator.

A fourth issue is the accuracy of confidence intervals computed using estimated quantiles of the pseudo-posterior. We find mixed results for confidence interval coverage: in

some cases coverage is very accurate, while in others the confidence intervals are too broad, so true size is smaller than the nominal size. Because the findings are mixed, we do not present tabular results, and we leave this issue for future research. It is perfectly feasible to use other means (asymptotic, bootstrap, Monte Carlo) of computing confidence intervals and standard errors for the SBIL estimator. For example, Li (2010) found that bootstrap confidence intervals are very accurate for the II estimator of the structural auction model discussed below. The same method could be used for the SBIL estimator. We do not pursue the issue further in this paper.

To implement the SMIL and SBIL estimators, we use between $S = 10^6$ and $S = 10^7$ simulated points drawn randomly from the parameter space, depending on the application. The auxiliary statistics we use are in most cases computationally inexpensive, so generating a large number of replications is not burdensome. The exceptions are the DSGE models, which requires approximately two days of time on a 32 core cluster per 10^6 replications of the auxiliary statistic[3]. For all problems, we use at least 5000 Monte Carlo replications at each design point. The nonparametric fit is done using the k nearest neighbors approach[4], using the ANN library (Arya, Malamatos and Mount, 2009; http://www.cs.umd.edu/~mount/ANN/). Using this C++ library, the KNN nonparametric fitting step requires at most several minutes of time on a single core. It is also a simple matter to switch to using approximate nearest neighbors, which can speed up the nonparametric fitting step if one uses an extremely large number of simulations. The number of neighbors k used for the nonparametric fit is chosen (with one exception) as $k = 1.5 \times S^{0.25}$, rounded down to the nearest integer. More careful choice using methods such as cross validation might improve the results, but we do not explore this possibility in this paper. We report the SBIL estimator computed as the posterior mean. The version computed as the posterior median gives very similar results. 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. For some of the applications (the MA and dynamic panel data models), prior beliefs such as stationarity or invertibility lead directly to the specification of at least some of the bounds of parameter space. For others (the auction model and the DSGE models) we have less information available regarding plausible bounds on at least some of the parameters. The issue of setting the parameter space in such cases is addressed in the subsection presenting the auction model.

7.1. **Dynamic panel data.** Gouriéroux, Phillips and Yu (2010; henceforth GPY) investigate the performance of the II estimator using a linear dynamic panel model

$$(17) \qquad\qquad y_{it} = \alpha_i + \phi_0 y_{it-1} + \epsilon_{it}$$

---

[3]Performing Monte Carlo on a cluster is quite straightforward. We use PelicanHPC (http://pelicanhpc.org/), a framework very similar to that described in Creel (2007).

[4]We also have used kernel regression, which gives very similar results to the KNN results reported here.

where $\epsilon_{it} \sim N(0,1)$, $\alpha_i \sim N(0,1)$, $\phi_0 = 0, 0.3, 0.6, 0.9$ and α_i and ϵ_i are independently distributed. The initial condition is

$$y_{i0}|\alpha_i \sim N\left(\frac{\alpha_i}{1-\phi_0}, \frac{1}{1-\phi_0^2}\right).$$

GPY use the (inconsistent) ML "fixed effects" estimator as the auxiliary statistic. They find that the II estimator outperforms a number of alternative estimators, in terms of root mean squared error (RMSE). While our asymptotic results do not straightforwardly generalize to dynamic panel data models (where the theory normally requires the number of time periods, T, to grow with sample size), we conjecture that the higher-order efficiency results also hold in this context. We here investigate this claim by comparing the performance of the SMIL and SBIL estimators to the II results obtained by GPY. GPY also report results for other bias correction methods such as jackknife and analytical bias correction and find that their II estimator dominates those; we therefore focus on the II estimator and do not reproduce the results for the other estimators. The parameter space is set to the stationary region $\phi_0 \in (-1, 1)$. We consider two auxiliary statistics: the same ML estimator as used by GPY, and also the ML estimator augmented with the OLS estimator of the naive model $y_{it} = \delta y_{it-1} + v_{it}$ that ignores the presence of individual effects.

The SMIL estimator requires a nonparametric density fit embedded inside an optimization problem, while the SBIL estimator eliminates the optimization. In the present case, the parameter to estimate is a scalar, so for this problem it is relatively easy to apply both the SMIL and SBIL estimators. By comparing the two in this relatively simple case, we can get an indication of whether focusing on the SBIL estimator in more computationally demanding cases is warranted by a comparable performance of the two estimators. To implement the SMIL, we use a different approach than what is outlined in equations (4) and (5). The reason for this to take advantage of the large set of replications of (θ^s, Z_n^s) that are already available after computing the SBIL estimator. Instead of operating on the conditional density $f_n(Z_n|\theta)$, we work with the joint density $f_n(Z_n, \theta)$. When θ^s is drawn from a uniform density, as is the case here, $f_n(Z_n, \theta)$ and $f_n(Z_n|\theta)$ are maximized at the same value of θ, because the marginal density of θ does not depend upon θ. We of course do not know the joint density, so it must be fit nonparametrically. We use the simple KNN density estimator given in equation 14.2 of Li and Racine (2007) to fit $f_n(Z_n, \theta)$. This nonparametric fit to the joint density, $\hat{f}_n(Z_n, \theta)$ is then maximized with respect to θ using a grid search, in order to deal with the rough, nondifferentiable nature of the KNN density estimator. Because θ is a scalar in the present case, use of grid search does not present a significant computational burden.

Table 1 presents the bias of the estimators, and Table 2 presents the root mean squared errors (RMSEs). In these Tables, the columns labeled II, SBIL and SMIL all refer to use of the auxiliary statistic $Z_n = \hat{\phi}_{ML}$, while the columns labeled SBIL(OI) and SMIL(OI) refer to use of the overidentifying auxiliary statistic $Z_n = \left(\hat{\phi}_{ML}, \hat{\delta}_{OLS}\right)$. Results for the inconsistent ML estimator are also presented, for reference. We see that the II and SBIL estimators have very small biases in almost all cases. With an exactly identifying auxiliary statistic, the estimators (except ML) all have similar biases and RMSEs, especially for

larger sample sizes. For small sample sizes, the SBIL estimator performs somewhat better than the II estimator, overall. When the difference favors the II estimator, it is small, but when it favors the SBIL estimator, it is larger. For the SMIL and SBIL estimators, it is easy to use an overidentifying auxiliary statistic, because no covariance matrix need be estimated. Looking at the columns labeled SMIL(OI) and SBIL(OI), we see that there are gains from doing so: bias is essentially unchanged, but RMSE is reduced considerably, especially for smaller sample sizes. There seems to be no reason to prefer SMIL to SBIL, as the RMSEs of the two are essentially the same in the case of the exactly identifying auxiliary statistic, while SBIL almost uniformly dominates SMIL when the overidentifying auxiliary statistic is used.

Based on the good performance of SBIL compared to SMIL in this example, and the fact that the two estimators are first order equivalent, we focus on SBIL in the remaining examples. Most of the remaining examples have parameter vectors of higher dimension, which would make a global maximization strategy such as grid search or simulated annealing more tedious to employ (recall that a nonparametric density fit must be done for each trial parameter value). The SBIL estimator does not require this optimization step, so it avoids this difficulty.

7.2. **Moving average.** The previous section compared the proposed estimators to a just identified II estimator. It is also desirable to compare to an overidentified II estimator, because this is the situation where it is necessary to estimate the efficient weight matrix in order to obtain an efficient II estimator, given the chosen auxiliary statistic. We would like to see if the SBIL estimator benefits from the fact that it does not require estimation of the efficient weight matrix. The first order moving average (MA(1)) model has been widely used to investigate the performance of the indirect inference estimator, and a *pth*-order autoregressive model is often used to generate the auxiliary statistic (see, for example, Gouriéroux, Monfort and Renault, 1993; Chumacero, 2001). In this section we estimate the MA(1) model

$$
\begin{aligned}
y_t &= \epsilon_t + \psi \epsilon_{t-1} \\
\epsilon_t &\sim i.i.d. \, N(0, \sigma^2)
\end{aligned}
$$

using sample sizes of n=50, 100 and 200 observations. The parameter ψ is one of the values $\{-0.95, -0.9, -0.5, 0, 0.5, 0.9, 0.95\}$, so the model is always invertible. The parameter σ is always equal to 1. The parameter vector is $\theta = (\psi, \sigma)$. We set the parameter space to $\Theta = (-1, 1) \times (0, 2)$, which imposes invertibility, which is needed for the parameter to be identified. The statistic Z_n is the vector of estimated parameters $(\rho_0, \rho_1, ..., \rho_P, \sigma_v^2)$ of an AR(P) model $y_t = \rho_0 + \sum_{p=1}^{P} \rho_p y_{t-p} + v_t$, fit to the data using ordinary least squares. For simplicity, we hold the order of the AR(P) model constant at $P = 10$ across the Monte Carlo replications. Thus, the dimension of Z_n is 12, while the dimension of θ is 2, so we have considerable overidentification.

We estimate θ using SBIL and II, where both are based on the auxiliary statistic defined in the last paragraph. The II estimator is computed using continuously updated

GMM (Hanson, Heaton and Yaron, 1996). The moment conditions that define the continuously updated indirect inference (CU-II) estimator are $m_n(\theta) = Z_n - \bar{Z}_{S,n}(\theta)$ where $\bar{Z}_{S,n}(\theta) = \frac{1}{S} \sum_{s=1}^{S} Z_n^s(\theta)$, and the weight matrix at each iteration is the inverse of $\Omega_n^S(\theta) = \frac{1}{S} \sum_{s=1}^{S} [Z_n^s(\theta) - \bar{Z}_{S,n}(\theta)] [Z_n^s(\theta) - \bar{Z}_{S,n}(\theta)]'$, where $S = 100$. For reference, we also estimate θ using the conditional maximum likelihood estimator (Gaussian MLE with ϵ_0 set to zero). When a replication of the ML or CU-II estimator lies in the non-invertible part of the parameter space, we use the observationally equivalent invertible parameter value in its place. The need for doing this and the means of doing so are explained by Chumacero (2001).

Table 3 reports the results. In this Table, SBIL(AR) refers to the SBIL estimator that uses the AR(10) auxiliary statistic, while SBIL(ML) is the SBIL estimator that uses the ML estimator as the auxiliary statistic. We see that the SBIL(AR) and CU-II estimators have biases that are of comparable magnitudes, overall. Comparing RMSEs, the SBIL(AR) estimator performs better than the CU-II estimator, almost uniformly. This result is not unexpected, given the previous theoretical results for higher order efficiency of MIL compared to CU-II. These theoretical grounds for efficiency plus the avoidance of estimation of the weight matrix appear to lead to real small sample efficiency gains. Comparing to the ML estimator, for the smaller sample size, SBIL(AR) has a larger RMSE than does ML, which is no doubt an indication that an AR(10) auxiliary model is excessively parameterized when the sample size is only 50. When the sample size is 200, the SBIL(AR) estimator has bias and RMSE comparable to those of the ML estimator. When the ML estimator is used as the auxiliary statistic for SBIL, there is no benefit in terms of RMSE when the sample size is 50, but for samples of size 100 and 200, the SBIL(ML) estimator has an RMSE lower than that of the ML estimator.

7.3. **Nonlinear panel model.** Section 7.1 explores a linear panel data model with normally distributed errors. One might expect that a nonlinear model could lead to a larger difference between the SBIL and II estimators, especially for smaller sample sizes, as in such a case the small sample distribution of the auxiliary statistic, which characterizes the objective functions of the IL estimators, could be less well approximated by the corresponding normal limiting distribution, which characterizes the objective function of the II estimator. To investigate this conjecture, we use the static logit panel model that Arellano and Bonhomme (2009) used in some of their Monte Carlo work to compare a set of semi-parametric nonlinear panel data estimators. Their static logit Monte Carlo design (see their Section 7.1) is used here to compare the SBIL and CU-II estimators. The design of the experiment is

$$y_{it} = \mathbf{1}\left[x_{it}\phi_0 + \alpha_{i0} + \epsilon_{it} > 0\right]$$

where $x_{it} \sim N(0,1)$ and the individual effects $\alpha_{i0} \sim N(\bar{x}_i, 1)$, where $\bar{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_{it}$. The ϵ_{it} are independent draws from the logistic CDF. The true value of $\phi_0 = 1$. We set $N \in \{30, 100\}$ and $T = 5$. The first component of the auxiliary statistic is the estimator of the misspecified logit model that results from the above model, with the exception that, erroneously, it is assumed that the individual effects are all identical. To be precise, it is the quasi-ML estimator resulting from logit estimation of the misspecified model $y_{it} =$

$1 [\alpha + x_{it}\phi + \epsilon_{it} > 0]$. The second component of the auxiliary statistic is the OLS estimator of the linear probability model $y_{it} = \alpha + x_{it}\phi + \eta_{it}$. The logit and OLS estimators of α and ϕ together yield an auxiliary statistic of dimension 4, so we have overidentification for the estimator of the scalar ϕ_0. We use SBIL and CU-II to estimate ϕ_0, using this auxiliary statistic. SBIL uses 2×10^6 simulations, and CU-II was implemented as described in the previous section. For both SBIL and CU-II, the parameter space for ϕ is set to $[0, 2]$ and the pseudo prior for SBIL is a uniform distribution over the parameter space.

Table 4 presents the results for bias, RMSE and mean absolute error (MAE). For both sample sizes, the SBIL estimator is less biased and has smaller RMSE and MAE than the II estimator. For the smaller sample size, the RMSE of the SBIL estimator is 88.4% that of the CU-II estimator, while for the larger sample size the percentage is 91.6%. This result supports the conjecture that the SBIL estimator will have better small sample performance than that of GMM-type estimators based on the same auxiliary statistic. Comparing these results to those for the linear dynamic panel data model, it seems that the nonlinearity of the model also contributes to accentuate the difference in performance of the SBIL and GMM-type estimators. For the sample size $N = 100, T = 5$, the MAE and bias results may be compared with the first panel of Table I in Arellano and Bonhomme (2009). Both SBIL and CU-II have less bias and lower MAE than any of the estimators considered by Arellano and Bonhomme. This is to be expected, because those estimators are semi-parametric, in that the distribution of the individual effects is unknown. The SBIL and CU-II estimators, in contrast, are based on simulations that require knowledge of the distribution of the fixed effects. The assumption that the distribution of the individual effects be known is quite implausible in this example. Nevertheless, the example serves to illustrate how the SBIL and II estimators can achieve a good bias reduction in small samples, through use of a simple naive auxiliary model, when one is able to write a fully simulable model.

7.4. **Structural model of an auction.** Li (2010) proposes to use indirect inference 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. The number of bidders is fixed at $N = 6$, and the sample size is $n = 100$, meaning that the outcomes of 100 auctions are observed. At each auction $i = 1, 2, ..., 100$, the quality, x_i, of the item being auctioned is the square of a uniform $(0, 2)$ random variable, to introduce heterogeneity in the values of the objects across the auctions. The 6 bidders draw their independent private values from a common exponential distribution with density

$$f(v|x_i) = \frac{1}{\exp(\theta_0 + \theta_1 x_i)} \exp\left(-\frac{v}{\exp(\theta_0 + \theta_1 x_i)}\right)$$

so that $\exp(\theta_0 + \theta_1 x_i)$ is the mean valuation of the item, over the bidders. The equilibrium strategy for the winning bid is

$$b_i^* = v_i^* - \frac{1}{F^{N-1}(v_i^*|x_i)} \int_0^{v_i^*} F^{N-1}(u|x_i) du$$

where v_i^* is the highest private valuation, and $F(\cdot|x_i)$ is the exponential distribution function. For a given value of N (6 in this case), symbolic computation software can be used to obtain an analytic solution for the winning bid, which facilitates simulation of the model. The observed data are the 100 values of $\{x_i, b_i^*\}$, and we seek to estimate θ_0 and θ_1. The true values are set to $\theta_0 = 1$ and $\theta_1 = 0.5$. Li presents results for indirect inference using two auxiliary statistics: the fitted coefficients of a pseudo ML estimator, and the OLS regression coefficients $(\widehat{\beta}_0, \widehat{\beta}_1)$ obtained by fitting the model $b_i^* = \beta_0 + \beta_1 x_i + \sigma \epsilon_i$.

To apply the SBIL estimator, we must specify the parameter space. The present application is interesting, because we have no clear *a priori* bounds for the two parameters θ_0 and θ_1. Outside of the Monte Carlo context, one would only have the sample data, but would not know the true parameter value. We discuss the issue of how the parameter space may be specified at some length, because it is a necessary step to apply the SBIL estimator. Our proposal is to start with a parameter space that seems conservatively large, and to check that it in fact contains elements that can generate simulated statistics Z_n^s that differ in important respects from the Z_n generated by the sample data. To do this, one can generate a preliminary set of Z_n^s setting S small enough to be convenient. Then one may compute the distance between each simulated statistic and the statistic using the sample data, giving the S distances d^s. Then one can sort the S replications of (θ^s, Z_n^s, d^s) by d^s and check that the θ^s that generate relatively small distances are always comfortably far away from the bounds of the proposed parameter space. If this is not the case, the parameter space can be expanded, and the procedure repeated again. Conversely, one may find evidence that the proposed parameter space is excessively broad, in that regions of the parameter space never generate statistics close to Z_n. Such simulations will not contribute to the nearest neighbors version of SBIL, and as such are wasted. This could be avoided by using importance sampling, but we here for simplicity take a brute force approach and simply choose initially a large parameter space and a moderate number of simulations, S for an intial exploration of the distribution of the statistic across different parameter values. We then shrink the parameter space removing parts whith little or no contribution to the posterior distribution.

We initially set the parameter space to $\Theta = (-5, 5) \times (0, 5)$. We generate a single sample at the true parameter value, and a fairly small number (10^5) simulated samples from the proposed parameter space. Inspection of the distribution of the auxiliary statistic used by Li reveals that the auxiliary statistic when sampling from the proposed parameter space presents some extreme outliers. This is a problem that may not be detected when using the II estimator with a limited number of replications of the auxiliary statistics (Li uses only one draw), because the II estimator maintains the underlying random draws fixed over the iterations, to avoid the phenomenon of "chatter" when doing the minimization to compute the estimator. The chances of encountering an outlying value of the auxiliary statistic are small, because only rare random draws generate outliers, by definition, and a fairly small number of draws are used. However, when a large number of auxiliary statistics are generated, as is the case with the SBIL estimator, outliers will eventually appear if the distribution of the auxiliary statistic has outliers in its support.

To address this, one can choose an auxiliary statistic that does not present outliers. Implementing this idea, we change the auxiliary model to $\log b_i^* = \beta_0 + \beta_1 x_i + \sigma \epsilon_i$, and consider two versions of the auxiliary statistic: $Z_n = (\widehat{\beta}_0, \widehat{\beta}_1)$ and $Z_n = (\widehat{\beta}_0, \widehat{\beta}_1, \log \widehat{\sigma})$. Inspection of the distribution of these auxiliary statistics reveals that they do not suffer from the presence of outliers. The distributions of the elements of the auxiliary statistics are notably non-Gaussian, however. For example, the distribution of $\log \widehat{\sigma}$ is bimodal. To refine the initial parameter space, we use the method suggested in the last paragraph. Of the 10^5 simulations, we use the minimal and maximal values of the θ^s that result in the 5000 Z_n^s closest to the single Z_n generated at the true parameter value, a procedure that is feasible when using real data. Note that the chosen rule for selecting the number of neighbors to use for estimation, $k = 1.5 \times S^{0.25}$, rounded downward, results in $k = 26$, so choosing the 5000th neighbor as the bound is quite conservative, as long as the distribution of Z_n^s changes sufficiently rapidly as θ^s changes. This last condition may be verified by inspection, at least when the dimension of the auxiliary statistic is small. The refined parameter space is $(\theta_0, \theta_1) \in \Theta = (-0.05, 2.40) \times (0.00, 1.95)$. After this initial exploration to set the auxiliary statistic and the parameter space, we increase S to 5×10^6, and proceed as normal.

Table 5 contains the results. SBIL (OLS) uses the just identifying auxiliary statistic which is the same as the second of Li's choices, except for the logarithmic transformation. Comparing to Li's results for the II estimator, the SBIL estimator is less biased than either of the II estimators, for both parameters, with a lower RMSE as well. RMSE is very much lower for the θ_0 parameter. Using the overidentifying auxiliary statistic (the entry labeled SBIL (extended OLS) we see that even better results obtain.

7.5. **Dynamic stochastic general equilibrium models.** In this subsection, we estimate two simple dynamic stochastic general equilibrium (DSGE) models. An and Schorfheide (2006) and Karagedikli *et al.* (2010) offer recent discussions of econometric methods for DSGE models, focusing on Bayesian estimation methods using Kalman or particle filtering and Markov chain Monte Carlo (MCMC). Winschel and Krätzig (2010) discuss recent advances in these areas. The SBIL estimator is similar to such methods in that it is a (pseudo) posterior mean or median, but there are some notable differences: the pseudo prior does not necessarily reflect beliefs, an auxiliary statistic plays an intermediate role, and simulation and nonparametric fitting replace filtering and MCMC. It is worth noting that the SMIL and SBIL estimators can accommodate nonlinearities and/or non-normal shocks in the model without any particular difficulties. To be able to estimate a DSGE model using SMIL or SBIL, the only requirement is that the model can be solved, by any appropriate means, and then simulated. Because successful nonparametric fitting requires a large number of simulations, we use a third-order perturbation solution, which combines good accuracy with moderate computational demands (Aruoba, Fernández-Villaverde and Rubió-Ramírez, 2006).

7.5.1. *A fully observed model with monopolistic competition.* The first model is a simple real business cycle model with monopolistic competition that was contributed as an example

by Fernández-Villaverde to the Dynare[5] web site. The model is explained in some detail in the Dynare User Guide, Chapter 3 (Mancini, 2010), which also gives full details of how Dynare can be used to solve and estimate the model using Bayesian MCMC methods. To facilitate comparison of methods, we use exactly the same model and parameter values as Mancini (2010, Chapter 3)[6].

The model is as follows: Households maximize expected discounted utility

$$E_t \sum_{s=0}^{\infty} \beta^s \left[\log c_{t+s} + \psi \log \left(1 - l_{t+s} \right) \right]$$

subject to the budget constraint and the accumulation of capital

(18) $$c_t + k_{t+1} = w_t l_t + r_t k_t + (1 - \delta) k_t$$

The variables are: c consumption; k capital; l labor; w real wages; r real price of capital. Production of an intermediate good y_{it} is done only by firm i of a continuum of firms between 0 and 1, and is given by a constant returns to scale production function

(19) $$y_{it} = k_{it}^{\alpha} \left(e^{z_t} l_t \right)^{1-\alpha}$$

Technology shocks z_t follow an AR(1) process:

(20) $$z_t = \rho z_{t-1} + \sigma \epsilon_t$$

where $\epsilon_t \sim IIN(0, 1)$. A final good producer has a constant elasticity of substitution production function

$$y_t = \left(\int_0^1 y_{it}^{\frac{\epsilon-1}{\epsilon}} \, di \right)^{\frac{\epsilon}{\epsilon-1}}$$

that aggregates intermediate goods into a final good demanded by consumers. The parameters of the model are α (technology); β (discount rate); δ (depreciation rate); ψ (consumption-leisure elasticity of substitution); ρ (AR1 parameter, technology shocks); σ (standard error, technology shocks); and ϵ (intermediate good elasticity of substitution).

The lower and upper bounds of the parameter space and the true parameter values are given in Table 6. The chosen limits are intended to be broad, in comparison to the fairly strongly informative priors that are often used when estimating DSGE models. They are also chosen so that the pseudo-prior mean is biased for the true parameter value, to illustrate the SBIL estimator's ability to recover from this bias. Our pseudo-prior $\pi(\theta)$ is a uniform distribution over the hypercube defined by the bounds of the parameter space.

Given a draw θ^s from the parameter space, first, the model is solved using Dynare, using a third order perturbation about the steady state. Once the model is solved, a simulation of length 180 is done, initialized at the steady state. We drop 100 observations, retaining the last 80 observations, which mimic 20 years of quarterly data. The observable

---

[5]Dynare (http://www.dynare.org/) is free software for solution of DSGE model using perturbation methods, and for estimation of such models using Bayesian MCMC methods, as well as maximum likelihood.

[6]Our Dynare code is a simple modification of Fernández-Villaverde's file rbc_monopolistic.mod, contained in the archive rbc.zip, which is available at http://www.dynare.org/documentation-and-support/examples/rbc.zip. Our modifications read true parameter values from a disk file, and provide analytic steady state values conditional on the parameter values.

variables are c, k, l, w, r, and y, in line with much empirical work (Guerron, 2010). With the 80 observations, we compute the auxiliary statistic Z_n^s. The elements of the auxiliary statistic are chosen with an eye to their ability to identify the parameters of the model. An advantage of having a fully structural model is that it is often suggestive of simple naive statistics that can identify the parameters of the model. For example, equation 18 can be solved for δ, then averaged over the data to get an estimate of δ. Given an estimate of δ, the sample average of the Euler equation

$$\frac{1}{c_t} = \beta E_t \left[\frac{1}{c_{t+1}} \left(1 + r_{t+1} - \delta \right) \right]$$

can be used to compute an estimate of β, if one simply ignores the expectation operator. Perhaps the parameter which is less obvious to identify is ϵ. This parameter is related to the marginal cost of production of the final good through $mc_t = (\epsilon - 1)/\epsilon$ (Mancini, 2010). We run a linear regression on the equation $w_t l_t + r_t k_t = a w_t + b k_t + c q_t + \eta_t$ and use the estimator $\hat{c}$ as a proxy for marginal cost. Then we compute a naive statistic related to ϵ as $\hat{\epsilon} = (1 - \hat{c})$. We do not use the inverse of this, as might seem more logical, because the inverse generates outliers when $\hat{c}$ is close to 1. As discussed above, one should avoid auxiliary statistics that present outliers. The auxiliary statistic has 13 elements, while the parameter vector has 7 elements, so we have a good deal of overidentification. The complete details of the vector of auxiliary parameters are given in the code that accompanies the paper.

The Bayesian approach to estimation of DGSE's that uses MCMC and filtering faces the issue of stochastic singularity (Ruge-Murcia, 2007), which means that the number of observable variables used to form the likelihood function is limited by the number of stochastic shocks to the model. This introduces the problem of selecting which variables to use, and this choice can have important effects on the estimation results (Ruge-Murcia, 2007; Guerron, 2010). The SBIL estimator does not face this problem: the entire set of observable variables can be used to compute the auxiliary statistic. In our example, there is a single stochastic shock to the model, but we use all of the endogenous variables to compute the elements of the auxiliary statistic.

The results are given in Table 6 and Figure 1. We can see that the SBIL estimator has a low bias for all of the parameters, and that the bias of the SBIL estimator is considerably lower than the bias of the pseudo-prior mean. The most notable bias is that for estimation of ρ, which is downward. The root mean square error of the SBIL estimator is also considerably lower than that of the pseudo-prior mean, for all parameters. We see that the density of the SBIL estimator moves toward and concentrates about the true parameter values, compared to the pseudo-prior distribution.

7.5.2. *A partially observed model with habit formation.* In real applications, it may be the case that some endogenous variables are not observable, with the capital stock being a leading example. The tailor-made auxiliary statistic of the previous DSGE example is infeasible if capital is not observed. A second limitation of the first example is the use of a very simple utility function: log utility. To address these two issues, and simply to provide more results for estimation of DSGE models, we consider another example,

taken from Ruge-Murcia (2010), which features a time non-separable utility function that has a curvature parameter.

Households maximize expected discounted utility

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

subject to the budget constraint and the accumulation of capital

$$c_t + k_{t+1} = y_t + (1 - \delta)k_t$$

where output and shocks are as in the previous example (respectively, equation 19, eliminating the subindex i, and equation 20). Utility depends on the curvature parameter γ and the habit formation parameter η. We use two designs. For the first, the true parameter values are: $\alpha = 0.36$, $\beta = 0.95$, $\delta = 0.025$, $\eta = 0.2$, $\gamma = 2$, $\rho = 0.85$, $\sigma = 0.04$, and $\psi = 3.197$. The second design sets $\eta = 0.4$, $\gamma = 4$ and $\psi = 13.562$, with the other parameters taking the same values as in the first design. Following Ruge-Murcia (2010), the value of ψ is set to make the steady state number of hours worked be 1/3 of the time endowment, given the other true parameter values.

The observable variables are consumption c_t, output y_t, and labor l_t, and the sample size is $n = 160$ (simulating 40 years of quarterly data). The capital stock is not observed, and this prevents the use of the tailor-made auxiliary statistic of the previous example. In the present case, the auxiliary statistic incorporates some quantities targeted to help identify specific parameters (for example, we expect that the average of c_t/y_t will help to identify γ), as well as means, variances, covariances and autocovariances of the observed variables. Because we are forced to use less targeted statistics, we use a larger number of them, to help to achieve precise estimation. The dimension of the auxiliary statistic we use is 33, more than twice as large as in the previous example. Because successful nonparametric fitting using KNN regression requires the number of neighbors to grow more slowly when the dimension of the conditioning variable increases, we modify our rule to become $k = 0.5 \times S^{0.25}$, rounded down to the nearest integer. We used 4×10^6 simulations to compute the SBIL estimator as a nonparametric conditional mean.

One last point is how additional prior information might be incorporated when estimating using SBIL. We assume that it is known that steady state hours are one third of the endowment. To use this information when estimating using SBIL, one can draw all parameters other than ψ from the chosen pseudo-prior. Then the value of ψ^s that leads to steady state hours equal to 1/3 is computed. All of the parameters together constitute θ^s. Thus, all trial θ^s incorporate the restriction. From this point on, SBIL is computed as usual. Our prior for all parameters other than ψ is uniform, with limits given in Table 7. Note that the prior means are biased for the true parameter values. For ψ, the marginal prior density is shaped like that of an exponentially distributed random variable. For the first design, the prior mean is considerably biased, while for the second it is less so.

Table 8 gives the results for the first design. We can see that the prior mean is biased for all parameters, while the SBIL estimator is an order of magnitude less so, with the exceptions of the parameters η and γ, where bias is reduced by about 50%. Looking at

RMSE, we see that the SBIL estimator successfully closes in on all of the true parameter values. The parameters that are estimated with less precision are those related to preferences: η, γ and ψ. The remaining parameters are estimated with very good precision. Table 7 presents the results for the second design. The results are qualitatively similar to those for the first design: the SBIL estimator continues to perform well even in the case of considerably stronger habit formation and risk aversion.

Ruge-Murcia also provides Monte Carlo results for estimation of this model, using SMM, with an auxiliary statistic similar to ours. Estimation by SMM is conceptually quite similar to estimation by SBIL, as was discussed above, and as shown by Ruge-Murcia, it is a another feasible method for estimation of a nonlinear DSGE with non-normal shocks. The possible advantages of SBIL compared to SMM are the avoidance of minimization (note that the SMM criterion may be nonconvex) and the avoidance of the need to estimate the efficient weight matrix. In spite of the use of the same model and a similar auxiliary statistic, our results are not directly comparable to Ruge-Murcia's. He fixes the parameters α and δ at their true values, citing evidence of weak identification in the context of other estimation methods applied to linearized models (Canova and Sala, 2009), and estimates the remaining parameters. Also, he sets the true value of the habit persistence parameter to $\eta = 0.8$, whereas we set $\eta \in \{0.2, 0.4\}$. We use these lower values of the habit persistence parameter because Dynare is occasionally not able to solve the model using a third order perturbation method when η takes on values close to 1. Our results show that α and δ are in fact well-identified using the SBIL estimator and our chosen auxiliary statistic, and we are able to estimate all of the model's parameters. Canova and Sala (2009) discuss possible weak identification of the parameters of linearized DSGE models when estimation is based on fitted impulse response functions from VARs. Our proposal is similar, in that one could use fitted impulse response functions as the statistic that defines the SBIL estimator. Our results are suggestive that inference though a statistic may not suffer such serious identification problems when the model is solved using a more accurate higher order solution method, rather than linearized. This is the case for the SBIL estimator, and it may hold for other estimators, as well. The issue is certainly worthy of additional investigation.

The SBIL estimator is quite simple to use for estimation of a DSGE model - one only needs to solve the model many times using different draws from the parameter space, and then compute an auxiliary statistic for each solution. This step takes time to perform, but it is very straightforward[7]. When this is completed, KNN regression is applied to compute the SBIL estimator. This second step only takes a minute or so. The only area where a researcher must use knowledge and judgment is in the choice of the auxiliary statistic. The availability of the structural model provides much useful guidance in this regard, as discussed above. 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

---

[7]This part of the problem is straightforward when the model can be solved reliably at any point in the parameter space. For our example, we have an analytic solution for the steady state of the model. With this, Dynare is able to solve the model without difficulty. More complicated models will normally be more difficult to solve, by whatever means is deemed appropriate. However, the problem of solving the model is shared by any estimation method.

considerably more complex MCMC/filtering combination that is currently widely used to estimate such models.

7.6. **Summary.** To summarize the Monte Carlo results, the SBIL estimator appears to give quite precise estimates for a variety of models, and in comparison to other estimators. Comparing to indirect inference (or CU-II), RMSE for the SBIL estimator is almost uniformly lower or equal to that of the II estimator, when the same auxiliary statistic is used (DPD, MA, nonlinear panel data, structural auction examples). In some cases, the difference in favor of the SBIL estimator is considerable, while for the few cases where II is favored the difference is small. The SBIL estimator may be conveniently applied using an overidentifying auxiliary statistic, because there is no need to estimate the efficient weight matrix (all examples). For the dynamic panel and structural auction models, we have seen that use of an overidentifying auxiliary statistic can lead to important efficiency gains in comparison to an exactly identified II estimator. The MA and nonlinear panel data examples show that an overidentified SBIL estimator may have RMSE considerably lower than that of an overidentified II estimator that uses the same auxiliary statistic. The dynamic panel data model provides evidence that agrees with our Proposition 3, which states that the SBIL and GMM-type estimators are first order equivalent when the GMM-type (in this case, the II estimator) uses an efficient weight matrix. In this example, the two estimators are essentially identical when larger samples are draws. The structural auction model shows that it may be necessary to take steps to control the presence of outliers by choosing the auxiliary statistic with some care. In that example we also addressed the issue of setting the parameter space. The DSGE examples shows that the SBIL estimator provides an interesting alternative to Bayesian methods that employ filtering and MCMC for estimation of macroeconomic models. In particular, we are able to successfully estimate all parameters of two simple models using moderately sized samples. Our method allows use of all observable variables for computation of the auxiliary statistic, and is not affected by the issue of stochastic singularity that introduces a variable selection problem when using the likelihood-based estimation using filtering and MCMC. Our method is also very simple to use.

8. Conclusions

This paper has introduced introduced indirect likelihood estimators to the econometric literature. We establish bias reduction and higher order efficiency properties, along with first order equivalence to well-known econometric estimators such as the simulated method of moments and indirect inference. We also provide quite extensive Monte Carlo examples that confirm that desirable theoretical properties manifest themselves in good finite sample performance for a variety of models. In particular, the SBIL estimator typically has a small bias, and a variance that is usually smaller than that of a comparable GMM-type estimator.

The proposed estimators are not in general fully asymptotically efficient, because the auxiliary statistic Z_n will not normally be a sufficient statistic. However, the possible loss of asymptotic efficiency does give some important benefits. The dimension reduction

achieved by working with a finite dimensional statistic rather than with the full sample converts potentially infinite dimensional problems (as the sample grows) into tractable finite dimensional problems. This is an important simplification when nonparametric estimation methods are used. Moreover, with a careful choice of auxiliary statistic, once can hope for approximate sufficiency. As we have seen in the DSGE examples, the SBIL estimator may be computed even when the auxiliary statistic is of fairly high dimension, at the cost of requiring more simulations. The possibility of using a fairly high (but finite) dimensional auxiliary statistic makes it reasonably hopeful that the statistic approximately spans the space of the efficient score, in which case the SBIL estimator will be approximately fully asymptotically efficient. Our Monte Carlo results for the dynamic panel and nonlinear panel examples can be compared to the results of other authors for other estimators, giving support to the good relative efficiency of the SBIL estimator. Additional support comes from our MA example, where the SBIL estimator often exhibits an RMSE smaller than that of the ML estimator.

The fact that the SBIL estimator may have better small sample performance than the ML estimator may be relevant when one seeks to estimate complex DSGE models. The combination of particle filtering and MCMC discussed above seeks to compute the ML estimator or related Bayesian likelihood-based estimators. The filtering/MCMC technology is relatively complicated to implement, and is computationally extremely demanding. In comparison, the SBIL estimator is simple to implement. In addition, it is certainly possible that the SBIL estimator could have better small sample performance than the ML estimator of such complex and often nonlinear models. An interesting avenue to explore would be to compare our estimator with the the MLE based on particle filtering/MCMC alternatives

In our implementation, we have focused on the basic sampler as given in equation (6) choosing the number of neighbors k through the simple rule $k = 1.5 \times S^{0.25}$. There is certainly scope for use of more sophisticated rules, such as cross-validation, or different kernels, which could lead to better performance. Similarly, more complicated samplers using importance sampling methods could be used to improve on the computation time. We leave these numerical issues for future research.

References

[1] Altonji, J. and L.M. Segal, 1996, "Small sample bias in GMM estimation of covariance structures," *Journal of Economic and Business Statistics* 14, 353-366.

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

[3] Andrews, D.W.K., 1993, "Exactly median-unbiased estimation of first order autoregressive/unit root models," *Econometrica* 61, 139–165.

[4] Arellano, M. and S. Bonhomme, 2009, "Robust priors in nonlinear panel data models," *Econometrica*, 77, 489-536.

[5] 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.

[6] Arya, S., T. Malamatos, and D.M. Mount, 2009, "Space-time tradeoffs for approximate nearest neighbor searching, *Journal of the ACM* 57, 1-54.

[7] Bahadur, R., S. Zabell, and J. Gupta, 1980, "Large deviations, tests, and estimates," in: I.M. Chaterabarli (Ed.), *Asymptotic Theory of Statistical Tests and Estimation*, pp. 33–64,. New York: Academic Press.

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

[9] Beaumont, M., J.-M. Cornuet, J.-M. Marin and C. Robert, 2009, "Adaptive approximate Bayesian computation", *Biometrika*, 96, 983-990.

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

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

[12] Bickel, P.J., F. Götze, and W.R. van Zwet, 1985, "A simple analysis of third-order efficiency of estimates," in L. Le Cam and R.A. Olshen (Eds.), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*. Wadsworth.

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

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

[15] Chumacero, R., 2001, "Estimating ARMA models efficiently", *Studies in Nonlinear Dynamics and Econometrics*, 5, 103-114.

[16] Collomb, G. and W. Härdle, 1986, "Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations," *Stochastic Processes and Their Applications* 23, 77-89.

[17] Creel, M., 2007, "I ran four million probits last night: HPC clustering with ParallelKnoppix," *Journal of Applied Econometrics*, 22, 215-223.

[18] Creel, M. and D. Kristensen, 2009, "Estimation of dynamic latent variable models using simulated nonparametric moments," UFAE and IAE Working Paper 792.09, http://ideas.repec.org/p/aub/autbar/792.09.html.

[19] Donald, S. and W.K. Newey, 2000, "A Jackknife Interpretation of the Continuous Updating Estimator," *Economics Letters* 67, 239-243.

[20] Doran, H.E. and P. Schmidt, 2006, "GMM estimators with improved finite sample properties using principal components of the weighting matrix, with an application to the dynamic panel data model," *Journal of Econometrics*, **133,** 387–409.

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

[22] Everaert, G. and L. Pozzi, 2007, "Bootstrap-based bias correction for dynamic panels", *Journal of Economic and Dynamics Control* 31, 1160-1184.

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

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

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

[26] Ghosh, J.K. (1994) *Higher Order Asymptotics*, Hayward: IMS.

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

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

[29] 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.

[30] Guerron, P., 2010, "What you match does matter: The effects of Data on DSGE Estimation," *Journal of Applied Econometrics* 25, 774-804.

[31] Gusev, S.I., 1975, "Asymptotic expansions associated with some statistical estimators in the smooth case I: Expansions of random variables," *Theory of Probability and Its Applications* 20, 470-498.

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

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

[34] Hahn, J. and G. Kuersteiner, 2002, "Asymptotically unbiased inference for a dynamic model with fixed effects when both n and T are large," *Econometrica* 70, 1639-1657.

[35] Hahn, J. and W.K. Newey, 2004, "Jackknife and analytical bias reduction for nonlinear panel models", *Econometrica* 72, 1295-1319.

[36] Hansen, L.P., J. Heaton and A. Yaron, 1996, "Finite-sample properties of some alternative GMM estimators", *Journal of Business and Economic Statistics* 14, 262-280.

[37] Horowitz, J. L., 1992, "A smoothed maximum score estimator for the binary response model," *Econometrica* 60, 505-531.

[38] Inoue, A. and M. Shintani, 2006, "Bootstrapping GMM estimators for time series," *Journal of Econometrics* 133, 531–555.

[39] Karagedikli, Ö., T. Matheson, C. Smith, C. and S.P. Vahey, 2010, "RBCs AND DSGEs: the computational approach to business cycle theory and evidence," *Journal of Economic Surveys*, 24, 113–136.

[40] Kezdi, G., J. Hahn and G. Solon, 2002, "Jackknife minimum distance estimation," *Economics Letters* 76, 35-45.

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

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

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

[44] Kristensen, D. and Y. Shin, 2008, "Estimation of dynamic models with nonparametric simulated maximum likelihood," CREATES Research Papers 2008-58, University of Aarhus.

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

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

[47] Mancini, T., 2010, "Dynare user guide: An introduction to the solution & estimation of DSGE models", http://www.dynare.org/documentation-and-support/user-guide/.

[48] Marjoram, P., J. Molitor, V. Plagnol and S. Tavaré, 2003, "Markov chain Monte Carlo without likelihoods", *Proceedings of the National Academy of Sciences, USA*, 100, 15324-15328.

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

[50] 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.

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

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

[53] 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.

[54] Rilstone, P., V.K. Srivastava and A. Ullah, 1996, "The second-order bias and mean squared error of nonlinear estimators," *Journal of Econometrics* 75, 369-395.

[55] Rothenberg, T.J. (1984), "Approximating the distributions of econometric estimators and test statistics," in: Z. Griliches and M.D. Intriligator (Eds.), *Handbook of Econometrics*, Vol. II, 881-935. Amsterdam: Elsevier Science.

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

[57] Ruge-Murcia, F., 2010, "Estimating nonlinear DSGE models by the simulated method of moments", working paper, Cahier 19-2010, CIREQ.

[58] Sisson, S., Y. Fan and M. Tanaka, 2007, "Sequential Monte Carlo without likelihoods", *Proceedings of the National Academy of Science, USA*, 104, 1760-1765.

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

[60] Skovgaard, I., 1986, "On multivariate Edgeworth expansions," *International Statistical Review* 54, 169-186.

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

[62] Tavaré, S., D. Balding, R. Griffiths and P. Donnelly, 1997, "Inferring coalescence times from DNA sequence data", *Genetics*, 145, 505-518.

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

[64] Zeitouni, O. and M. Gutman, 1991, "On universal hypothesis testing via large deviations," *IEE Transactions on Information Theory* 37, 285–290.

[65] Zhang, P., 1996, "Nonparametric Importance Sampling," *Journal of the American Statistical Association* 91, 1245-1253.

APPENDIX A: PROOFS

Proof. [Proposition 1] We first investigate the MIL estimator: To this end, first note that by Lemma 1 the log-likelihood satisfies

$$\frac{1}{n}\log f(Z_n|\theta) = \tfrac{1}{n}\log\phi_n^*(Z_n|\theta) + LR_n(\theta) = \frac{1}{n}\log\phi_n^*(Z_n|\theta) + o_P(1/\sqrt{n})$$

uniformly in θ. Thus, for the first-order analysis, we can treat $L_n(\theta) := \log\phi_n^*(Z_n|\theta)$ as the actual log-likelihood. To show consistency, note that uniformly in $\theta \in \Theta$:

$$
\begin{aligned}
\frac{1}{n}L_n(\theta) &= -\frac{1}{2n}\log(|\Omega(\theta)|) - \frac{T_n(\theta)' T_n(\theta)}{2n} + o_P(1) \\
&= -\frac{1}{2}(Z(\theta_0) - Z(\theta))'\Omega^{-1}(\theta)(Z(\theta_0) - Z(\theta)) + o_P(1) \\
&=: L(\theta) + o_P(1),
\end{aligned}
$$

(21)

where $L(\theta)$ is a continuous function 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 MLE is consistent.

Next, we show asymptotic normality: With $\dot{Z}^{(i)}(\theta) = \partial Z(\theta)/(\partial\theta_i)$ and $\dot{\Omega}^{(i)}(\theta) = \partial\Omega(\theta)/(\partial\theta_i)$,

$$
\begin{aligned}
\Delta_{n,i}(\theta) : &= \frac{\partial L_n(\theta)}{\partial\theta_i} \\
&= -\frac{1}{2}\Omega^{-1}(\theta)\dot{\Omega}^{(i)}(\theta) - \sqrt{n}T_n(\theta)'\Omega^{-1/2}(\theta)\dot{Z}^{(i)}(\theta) + \frac{1}{2}T_n(\theta)'\dot{\Omega}^{(i)}(\theta)T_n(\theta) \\
&= -\sqrt{n}T_n(\theta)'\Omega^{-1/2}(\theta)\dot{Z}^{(i)}(\theta) + o_P(\sqrt{n})
\end{aligned}
$$

and with $\ddot{Z}^{(i,j)}(\theta) = \partial^2 Z(\theta)/(\partial\theta_i\partial\theta_j)$ and $\ddot{\Omega}^{(i,j)}(\theta) = \partial^2\Omega(\theta)/(\partial\theta_i\partial\theta_j)$,

$$
\begin{aligned}
J_{n,ij}(\theta) : &= \frac{1}{n}\frac{\partial^2 L_n(\theta)}{\partial\theta_i\partial\theta_j} \\
&= \frac{1}{n}\Omega^{-2}(\theta)\dot{\Omega}^{(i)}(\theta)\dot{\Omega}^{(j)}(\theta) - \frac{1}{n}\Omega^{-1}(\theta)\ddot{\Omega}^{(i,j)}(\theta) \\
&\quad + \dot{Z}^{(i)}(\theta)'\Omega^{-1}(\theta)\dot{Z}^{(j)}(\theta) + T_n(\theta)'\Omega^{-1/2}(\theta)\ddot{Z}^{(i,j)}(\theta)/\sqrt{n} + o_P(1),
\end{aligned}
$$

With $J(\theta)$ defined in Assumption 3, it now holds that

(22)
$$\frac{1}{\sqrt{n}}\Delta_n(\theta_0) = -T_n(\theta_0)'\Omega^{-1/2}(\theta_0)\dot{Z}(\theta_0) + o_P(1) \to^d N(0, J(\theta_0)),$$

and, uniformly in θ, $J_n(\theta) = J(\theta) + o_P(1)$. Since the score of the log-likelihood converges weakly towards a normal distribution while the Hessian converges uniformly towards a non-singular limit in probability, it now follows by a standard Taylor expansion of the score that the MILE is $\sqrt{n}$-asymptotically normally distributed with asymptotic variance $J^{-1}(\theta_0)$.

Next, the properties of the BIL are established by verifying Assumptions 1-4 in Chernozhukov and Hong (2003), CH henceforth, with $L_n(\theta)$ chosen as above. First note that CH's Assumptions 1-2 are satisfied by our Assumption 1. 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. $\square$

Proof. [Proposition 2] This follows directly from Chernozhukov and Hong (2003, Theorem 3) since eqs. (22) and $J_n(\theta) = J(\theta) + o_P(1)$ imply that the generalized information equality holds. $\square$

Proof. [Proposition 3] By assumption, $\bar{Z}_n(\theta) = Z(\theta) + o(1/\sqrt{n})$, while $Z_n = Z_n(\theta_0) \rightarrow^P Z(\theta_0)$. Thus, $D_n(\theta) = D(\theta) + o_p(1)$, where the limit is given by

$$D(\theta) = \frac{1}{2}(Z(\theta_0) - Z(\theta))' \Omega^{-1}(\theta_0)(Z(\theta_0) - Z(\theta)).$$

By Assumption 3 in conjunction with standard arguments, it now follows that $\hat{\theta}_{GMM}$ is consistent. To derive its asymptotic distribution, first note that $\hat{\theta}_{GMM}$ solves

$$0 = \frac{\partial D_n(\theta)}{\partial \theta'} = -\frac{\partial \bar{Z}_n(\theta)}{\partial \theta}' W_n(Z_n - \bar{Z}_n(\theta)) = -\dot{Z}(\theta)' W_n(Z_n - Z(\theta)) + o_P(1/\sqrt{n}),$$

where, by Assumption 2,

$$Z_n - Z(\theta) = Z_n - Z(\theta_0) - \dot{Z}(\bar{\theta})(\theta - \theta_0),$$

where $\bar{\theta}$ lies on the line between θ and θ_0. Combining these two equations,

$$
\begin{aligned}
0 &= -\left\{\dot{Z}(\hat{\theta}_{GMM})\right\}' W_n(Z_n - Z(\hat{\theta}_{GMM})) + o_P(1/\sqrt{n}) \\
&= -\dot{Z}(\hat{\theta}_{GMM})' W_n\{Z_n - Z(\theta_0)\} + \dot{Z}(\hat{\theta}_{GMM})' W_n \dot{Z}(\bar{\theta})(\hat{\theta}_{GMM} - \theta_0) + o_P(1/\sqrt{n}).
\end{aligned}
$$

The result now follows by Assumption 2 together with $W_n \rightarrow^P \Omega^{-1}(\theta_0)$. $\square$

Proof. [Proposition 4] First, consider the two-step GMM estimator, $\hat{\theta}_{GMM}$. With

$$m_n(\theta) = (Z_n - \bar{Z}_n(\theta))' W_n \frac{\partial \bar{Z}_n(\theta)}{\partial \theta},$$

we can apply Lemma 2. The first and second order derivatives are given by

$$\frac{\partial m_n(\theta_0)}{\partial \theta} = -\frac{\partial \bar{Z}_n(\theta_0)}{\partial \theta}' W_n \frac{\partial \bar{Z}_n(\theta_0)}{\partial \theta} + (Z_n - \bar{Z}_n(\theta))' W_n \frac{\partial^2 \bar{Z}_n(\theta_0)}{\partial \theta^2},$$

and

$$\frac{\partial^2 m_n(\theta_0)}{\partial \theta^2} = -3\frac{\partial^2 \bar{Z}_n(\theta_0)}{\partial \theta^2}' W_n \frac{\partial \bar{Z}_n(\theta_0)}{\partial \theta} + (Z_n - \bar{Z}_n(\theta))' W_n \frac{\partial^3 \bar{Z}_n(\theta_0)}{\partial \theta^3}$$

With

$$
\begin{aligned}
D\bar{m}_n &= -\frac{\partial \bar{Z}_n(\theta_0)}{\partial \theta}' \Omega_n^{-1}(\theta_0) \frac{\partial \bar{Z}_n(\theta_0)}{\partial \theta}, \\
D^2\bar{m}_n &= -3\frac{\partial^2 \bar{Z}_n(\theta_0)}{\partial \theta^2} \Omega_n^{-1}(\theta_0) \frac{\partial \bar{Z}_n(\theta_0)}{\partial \theta},
\end{aligned}
$$

where $\Omega_n^{-1}(\theta_0)$ denotes the variance of $Z_n - \bar{Z}_n(\theta)$, and Δ_n defined in the proposition,

$$
\begin{aligned}
A_n : \; &= \; \frac{\partial m_n(\theta_0)}{\partial \theta} - D\bar{m}_n \\
&= \; \frac{\partial^2 \bar{Z}_n(\theta)'}{\partial \theta^2} \Omega_n^{-1}(\theta_0)(Z_n - \bar{Z}_n(\theta)) - \frac{\partial \bar{Z}_n(\theta_0)'}{\partial \theta} \Delta_n \frac{\partial \bar{Z}_n(\theta_0)}{\partial \theta} \\
&\quad + \frac{\partial^2 \bar{Z}_n(\theta)'}{\partial \theta^2} \Delta_n (Z_n - \bar{Z}_n(\theta)) \\
&= \; \frac{\partial^2 \bar{Z}_n(\theta)'}{\partial \theta^2} \Omega_n^{-1}(\theta_0)(Z_n - \bar{Z}_n(\theta)) - \frac{\partial \bar{Z}_n(\theta_0)'}{\partial \theta} \Delta_n \frac{\partial \bar{Z}_n(\theta_0)}{\partial \theta} + O_P(1/n).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
E[A_n m_n(\theta_0)] \; &= \; \frac{\partial^2 \bar{Z}_n(\theta)'}{\partial \theta^2} \Omega_n^{-1}(\theta_0) E\left[(Z_n - \bar{Z}_n(\theta))(Z_n - \bar{Z}_n(\theta))'\right] \Omega_n^{-1}(\theta_0) \frac{\partial \bar{Z}_n(\theta)}{\partial \theta} \\
&\quad - \frac{\partial \bar{Z}_n(\theta_0)'}{\partial \theta} E\left[\Delta_n \frac{\partial \bar{Z}_n(\theta_0)}{\partial \theta}(Z_n - \bar{Z}_n(\theta))'\right] \Omega_n^{-1}(\theta_0) \frac{\partial \bar{Z}_n(\theta)}{\partial \theta} \\
&= \; \frac{1}{n} \frac{\partial^2 \bar{Z}_n(\theta)'}{\partial \theta^2} \Omega_n^{-1}(\theta_0) \frac{\partial \bar{Z}_n(\theta)}{\partial \theta} \\
&\quad - \frac{1}{n} \frac{\partial \bar{Z}_n(\theta_0)'}{\partial \theta} E\left[\Delta_n \frac{\partial \bar{Z}_n(\theta_0)}{\partial \theta}(Z_n - \bar{Z}_n(\theta))'\right] \Omega_n^{-1}(\theta_0) \frac{\partial \bar{Z}_n(\theta)}{\partial \theta} \\
&\simeq \; \frac{1}{n}\left[\frac{1}{3} D^2 \bar{m} + B_{W,n}\right]
\end{aligned}
$$

with $B_{W,n}$ defined in the proposition. The other bias component can be written as:

$$
\begin{aligned}
E[m_n^2(\theta_0)] \; &= \; \frac{\partial \bar{Z}_n(\theta)'}{\partial \theta} \Omega_n^{-1}(\theta_0) E\left[(Z_n - \bar{Z}_n(\theta))(Z_n - \bar{Z}_n(\theta))'\right] \Omega_n^{-1}(\theta_0) \frac{\partial \bar{Z}_n(\theta)}{\partial \theta} \\
&= \frac{1}{n} \; \frac{\partial \bar{Z}_n(\theta)'}{\partial \theta} \Omega_n^{-1}(\theta_0) \frac{\partial \bar{Z}_n(\theta)}{\partial \theta} \simeq -\frac{1}{n} J(\theta_0).
\end{aligned}
$$

Thus, by Lemma 2,

$$
\begin{aligned}
E[\hat{\theta}_{GMM}] - \theta_0 \; &\simeq \; -J^{-2}(\theta_0)\left\{E[A_n m_n(\theta_0)] - \frac{1}{2}\frac{D^2\bar{m}_n}{D\bar{m}_n} E[m_n^2(\theta_0)]\right\}. \\
&\simeq \; \frac{1}{n} J^{-2}(\theta_0)\left\{\frac{1}{6} D^2\bar{m} + B_{W,n}\right\}
\end{aligned}
$$

Next, consider the CU estimator: It is easily checked that the expansion goes through with

$$
m_n(\theta) := 2\frac{\partial \bar{Z}_n(\theta)'}{\partial \theta} \Omega_n^{-1}(\theta)(Z_n - \bar{Z}_n(\theta)) + (Z_n - \bar{Z}_n(\theta))' \frac{\partial \Omega_n^{-1}(\theta)}{\partial \theta}(Z_n - \bar{Z}_n(\theta)),
$$

and $D\bar{m}_n$ and $D^2\bar{m}_n$ given as before. However, in the case of CU,

$$
A_n := \frac{\partial m_n(\theta_0)}{\partial \theta} - D\bar{m}_n = \frac{\partial^2 \bar{Z}_n(\theta)'}{\partial \theta^2} \Omega_n^{-1}(\theta_0)(Z_n - \bar{Z}_n(\theta)) + O_P(1/n)
$$

and so the bias term due to the first-step estimation of the weighting matrix vanishes and we obtain the claimed result.

Finally, consider the MIL estimator: Since $LR(\theta) = o_P(1/n^2)$, we can choose $m_n(\theta) = n^{-1}\partial \log f_n^*(Z_n|\theta) / (\partial \theta)$ such that

$$\frac{\partial m_n(\theta)}{\partial \theta} = \frac{1}{n}\frac{\partial^2 \log f_n^*(Z_n|\theta)}{\partial \theta^2}, \quad \frac{\partial^2 m_n(\theta)}{\partial \theta^2} = \frac{1}{n}\frac{\partial^3 \log f_n^*(Z_n|\theta)}{\partial \theta^3}.$$

From the definition of $f_n^*(Z_n|\theta)$, $m_n(\theta) = m_{n,1}(\theta) + m_{n,2}(\theta)$, where the first term is the Gaussian component,

$$m_{n,1}(\theta) \simeq \dot{Z}(\theta)' \Omega^{-1}(\theta)(Z_n - Z(\theta)),$$

while the second one is due to the higher-order component,

$$\begin{aligned} m_{n,2}(\theta) &\simeq \frac{1}{n^{3/2}}\frac{\partial \pi_1(T_n(\theta)|\theta)/\partial\theta}{1 + \pi_1(T_n(\theta)|\theta)/\sqrt{n}} \\ &\simeq \frac{1}{n^{3/2}}\frac{\partial \pi_1(T_n(\theta)|\theta)}{\partial\theta} \\ &\simeq -\frac{1}{n}\pi_1^{(1)}(T_n(\theta)|\theta)\,\Omega^{-1/2}(\theta)\dot{Z}(\theta). \end{aligned}$$

The derivatives satisfy

$$\frac{\partial m_{1,n}(\theta)}{\partial\theta} \simeq \frac{1}{2}\ddot{Z}(\theta)'\Omega^{-1}(\theta)(Z_n - Z(\theta)) - \dot{Z}(\theta)'\Omega^{-1}(\theta)\dot{Z}(\theta),$$

$$\begin{aligned} \frac{\partial m_{2,n}(\theta)}{\partial\theta} &\simeq -\frac{1}{n}\pi_1^{(1)}(T_n(\theta)|\theta)\,\Omega^{-1/2}(\theta)\ddot{Z}(\theta) \\ &\quad + \frac{1}{\sqrt{n}}\dot{Z}(\theta)'\Omega^{-1/2}(\theta)\pi_1^{(2)}(T_n(\theta)|\theta)\,\Omega^{-1/2}(\theta)\dot{Z}(\theta), \end{aligned}$$

and

$$\frac{\partial^2 m_{1,n}(\theta)}{\partial\theta^2} \simeq \dddot{Z}(\theta)'\Omega^{-1}(\theta)(Z_n - Z(\theta)) - 3\ddot{Z}(\theta)'\Omega^{-1}(\theta)\dot{Z}(\theta),$$

$$\begin{aligned} \frac{\partial^2 m_{2,n}(\theta)}{\partial\theta^2} &\simeq -\frac{1}{n}\pi_1^{(1)}(T_n(\theta)|\theta)\,\Omega^{-1/2}(\theta)\dddot{Z}(\theta) \\ &\quad + \frac{2}{\sqrt{n}}\dot{Z}(\theta)'\Omega^{-1/2}(\theta)\pi_1^{(2)}(T_n(\theta)|\theta)\,\Omega^{-1/2}(\theta)\dot{Z}(\theta) \\ &\quad + \sum_i \dot{Z}(\theta)'\Omega^{-1/2}(\theta)\tilde{\pi}_{1,i}^{(3)}(T_n(\theta)|\theta)\,\Omega^{-1/2}(\theta)\dot{Z}(\theta), \end{aligned}$$

where $\tilde{\pi}_{1,i}^{(3)}(T_n(\theta)|\theta) = \partial\pi_1^{(2)}(T_n(\theta)|\theta)/(\partial t_i)\dot{T}_{n,i}(\theta)/\sqrt{n}$. Since $\bar{T}_n(\theta_0) = O(1/\sqrt{n})$, we can choose $D\bar{m}_n$ and $D\bar{m}_n^2$ as for the GMM and CU estimators except that $Z(\theta)$ replaces $\bar{Z}_n(\theta)$. Next, in order to obtain an expression of the bias, we Taylor-expanding w.r.t. the statistic: With $\bar{Z}_{0,n} := \bar{Z}_n(\theta_0)$, $\bar{f}_n^* := \bar{f}_n^*(\bar{Z}_{0,n}|\theta_0)$ and $\bar{T}_n(\theta) := \sqrt{n}\Omega^{-1/2}(\bar{Z}_{0,n} - Z(\theta))$,

$$\frac{\partial^i m_n(\theta)}{\partial\theta^i} \simeq \frac{1}{n}\frac{\partial^{i+1}\log\bar{f}_n^*(\theta)}{\partial\theta^i} + \frac{1}{n}\frac{\partial^{i+2}\log\bar{f}_n^*(\theta)}{\partial\theta^i\partial z}(Z_n - \bar{Z}_{0,n}),$$

for $i = 0, 1, 2$, where

$$\frac{1}{n}\frac{\partial^2\log\bar{f}_n^*(\theta)}{\partial\theta\partial z} \simeq \dot{Z}(\theta)'\Omega^{-1}(\theta) + \frac{1}{\sqrt{n}}\Omega^{-1/2}(\theta)\pi_1^{(2)}(\bar{T}_n(\theta)|\theta)\,\Omega^{-1/2}(\theta)\dot{Z}(\theta),$$

$$\frac{1}{n}\frac{\partial^3 \log \bar{f}_n^*\left(\theta\right)}{\partial \theta^2 \partial z} \simeq \ddot{Z}\left(\theta\right)' \Omega^{-1}\left(\theta\right) - \frac{1}{\sqrt{n}}\Omega^{-1/2}\left(\theta\right)\pi_1^{(2)}\left(\bar{T}_n\left(\theta\right)|\theta\right)\Omega^{-1/2}\left(\theta\right)\ddot{Z}\left(\theta\right)$$
$$+\sum_i \dot{Z}\left(\theta\right)'\Omega^{-1/2}\left(\theta\right)\bar{\pi}_{1,i}^{(3)}\left(\bar{T}_n\left(\theta\right)|\theta\right)\Omega^{-1/2}\left(\theta\right)\dot{Z}\left(\theta\right),$$

where $\bar{\pi}_{1,i}^{(3)}\left(t\right)$ is defined in the proposition, and

$$\frac{1}{n}\frac{\partial^4 \log \bar{f}_n^*\left(\theta\right)}{\partial \theta^3 \partial z} \simeq \dddot{Z}\left(\theta\right)'\Omega^{-1}\left(\theta\right) + \frac{1}{\sqrt{n}}\Omega^{-1/2}\left(\theta\right)\pi_3^{(2)}\left(\bar{T}_n\left(\theta\right)|\theta\right)\Omega^{-1/2}\left(\theta\right)\dddot{Z}\left(\theta\right)$$
$$+2\sum_i \dot{Z}\left(\theta\right)'\Omega^{-1/2}\left(\theta\right)\bar{\pi}_{3,i}^{(3)}\left(\bar{T}_n\left(\theta\right)|\theta\right)\Omega^{-1/2}\left(\theta\right)\dot{Z}\left(\theta\right).$$

We note that $\bar{T}_n\left(\theta_0\right)=O\left(1/\sqrt{n}\right)$ such that $\pi_1^{(i)}\left(\bar{T}_n\left(\theta\right)\right)\simeq \pi_1^{(i)}\left(0\right)$. Thus,

$$E\left[A_n m_n\left(\theta_0\right)\right] \simeq \frac{1}{n^2}\frac{\partial^2 \log \bar{f}_n^*\left(\theta_0\right)}{\partial \theta \partial z'}E\left[\left(Z_n - \bar{Z}_n\left(\theta_0\right)\right)\left(Z_n - \bar{Z}_n\left(\theta_0\right)\right)'\right]\frac{\partial^3 \log \bar{f}_n^*\left(\theta_0\right)}{\partial \theta^2 \partial z}$$
$$\simeq \frac{1}{3n}D^2 \bar{m} + \frac{1}{n}B_\pi,$$

and

$$E\left[m_n^2\left(\theta_0\right)\right] \simeq \frac{1}{n^2}\frac{\partial^2 \log \bar{f}_n^*\left(\theta_0\right)}{\partial \theta \partial z'}E\left[\left(Z_n - \bar{Z}_n\left(\theta_0\right)\right)\left(Z_n - \bar{Z}_n\left(\theta_0\right)\right)'\right]\frac{\partial^2 \log \bar{f}_n^*\left(\theta_0\right)}{\partial \theta \partial z} \simeq \frac{1}{n}J\left(\theta_0\right).$$

Lemma 2 now yields the claimed result. $\qquad \square$

Proof. [Proposition 5] As usual, we can treat $f^*\left(Z_n|\theta\right)$ as the actual likelihood due to Lemma 1. By an rth order Taylor expansion of the corresponding score equation w.r.t. θ,

$$(23) \qquad 0 = W_{n,1}\left(Z_n\right) + \sum_{i=1}^{r}\frac{1}{i!}W_{n,i}\left(Z_n\right)\left(\hat{\theta}_{MIL} - \theta_0\right)^i + R_n, =: A\left(W_n\left(Z_n\right),\hat{\theta}_{MIL}\right) + R_n,$$

where $W_n\left(z\right) = \left(W_{n,1}\left(z\right),...,W_{n,r}\left(z\right)\right)$ with $W_{n,i}\left(z\right) = n^{-1}\partial^i \log f_n^*\left(z|\theta_0\right)/\left(\partial \theta_0^i\right)$, and

$$R_n = n^{-1}\left|\partial^r \log f_n^*\left(z|\theta\right)/\left(\partial \theta^r\right)\right|_{\theta=\bar{\theta}}\left(\hat{\theta}_{MIL} - \theta_0\right)^r.$$

First, ignore R_n and redefine $\hat{\theta}_{MIL}$ as the solution to $A\left(W_n\left(Z_n\right),\hat{\theta}_{MIL}\right) = 0$. From the expression of $f^*\left(Z_n|\theta\right)$, it is easily seen that

$$W_n\left(Z\left(\theta_0\right)\right) = W_\infty\left(Z\left(\theta_0\right)\right) + \sum_{i=1}^{r}\frac{1}{n^{i/2}}M_i + o\left(n^{-r/2}\right),$$

where M_i are constants depending on derivatives of the polynomials $\pi_1,...,\pi_r$ and $W_{\infty,i}\left(Z\left(\theta_0\right)\right)$ is the leading term of $n^{-1}\partial^i \log f^*\left(Z\left(\theta_0\right)|\theta_0\right)/\left(\partial \theta_0^i\right)$. In particular, the limiting score and Hessian satisfy $\bar{W}_{\infty,1}\left(Z\left(\theta_0\right)\right) = 0$ and $\bar{W}_{\infty,2}\left(Z\left(\theta_0\right)\right) = -J\left(\theta_0\right)$. Thus, $A\left(W_\infty\left(Z\left(\theta_0\right)\right),\theta_0\right) = 0$, and $\partial A\left(W_\infty\left(Z\left(\theta_0\right)\right),\theta\right)/\partial \theta|_{\theta=\theta_0} = -J\left(\theta_0\right)$ has full rank. Hence, by the implicit function theorem, there exists an analytic function $H\left(w\right)$ in a neighborhood of $W_\infty\left(Z\left(\theta_0\right)\right)$ such that $\theta_0 = H\left(W_\infty\left(Z\left(\theta_0\right)\right)\right)$. Moreover, for all n large enough, the solution $\theta_{0,n}$ to $A\left(W_n\left(Z\left(\theta_0\right)\right),\theta_{0,n}\right) = 0$, can be expressed as $\theta_{0,n} = H\left(W_n\left(Z\left(\theta_0\right)\right)\right)$ since $W_n\left(Z\left(\theta_0\right)\right)$

lies in a neighborhood of $W_\infty(Z(\theta_0))$ for all n large enough. The sequence $\theta_{0,n}$ satisfies

$$
\begin{aligned}
\theta_{0,n} - \theta_0 &= H(W_n(Z(\theta_0))) - H(W_\infty(Z(\theta_0))) \\
&= \sum_{i=1}^r \frac{\partial^i H(W_\infty(Z(\theta_0)))}{\partial w^i} [W_n(Z(\theta_0)) - W_\infty(Z(\theta_0))]^i + o\left(n^{-r/2}\right) \\
&=: \sum_{j=1}^r \frac{1}{n^{j/2}} \tilde{M}_j + o\left(n^{-r/2}\right),
\end{aligned}
$$

where $\tilde{M}_j$ is a constant depending on $M_1, ..., M_r$ and the first r derivatives of $H(W_\infty(Z(\theta_0)))$, $j = 1, ..., r$.

We obtain an Edgeworth expansion of $\hat{\theta}_{MIL} - \theta_{0,n} = H(W_n(Z_n)) - H(W_n(Z(\theta_0)))$ by applying the general result of Phillips (1977) for Edgeworth expansions of transformations of random sequences: We define the following sequence of functions

$$
e_n(q) := H(W_n(q + Z(\theta_0))) - H(W_n(Z(\theta_0))),
$$

such that $e_n := \hat{\theta}_{MIL} - \theta_{0,n} = e_n(q_n)$, where $q_n := Z_n - Z(\theta_0)$, and verify Phillips (1977, Assumptions 3-5): First, since the distribution of the normalized statistic $T_n(\theta_0) = \sqrt{n} q_n$ satisfies an Edgeworth expansion by Assumption 4, Phillips (1977, Assumption 3) holds. Next, the two function H and W_n are both r times continuously differentiable and the derivatives of $W_n(z)$ converges towards those of $W_\infty(z)$. Thus, $e_n(q)$ is r times differentiable with its derivatives uniformly bounded in a neighborhood around 0. Finally, we know from the implicit function theorem that $\partial H(W_\infty(Z(\theta_0)))/(\partial w)$ has full rank while it is easily checked that $\partial W_{\infty,1}(Z(\theta_0))/(\partial z) = \Omega^{-1/2}(\theta_0)\dot{Z}(\theta_0)$. Hence, by the chain rule, $|\partial e_n(q)/\partial q|$ is bounded away from zero as $n \to \infty$. This shows that Phillips (1977, Assumptions 4-5) hold.

We have shown that $\sqrt{n} e_n$ admits an Edgeworth expansion, say

$$
f_{e_n}^*(x) = \phi(x)\left[1 + \sum_{i=1}^r n^{-i/2} \bar{\pi}_i(x)\right].
$$

This in turn implies that the distribution of $\epsilon_n := \sqrt{n}(\hat{\theta}_{MIL} - \theta_0) = \sqrt{n} e_n + b_n$, where $b_n = \sqrt{n}(\theta_{0,n} - \theta_0) = \sum_{j=1}^r n^{-j/2} M_j + o(n^{-r/2})$, can be approximated by

$$
f_{\epsilon_n}^*(x) = \phi(x - b_n)\left[1 + \sum_{i=1}^r n^{-i/2} \bar{\pi}_i(x - b_n)\right].
$$

Expanding around $f_{e_n}^*(x)$ and rearranging terms, we then obtain the desired result where the coefficients of the polynomial $\tilde{\pi}_i(x)$ depend on the ones of $\bar{\pi}_j(x)$ and the coefficients $M_j, j = 1, ..., r$.

Finally, we have to verify that we are allowed to ignore the remainder term R_n in the Taylor expansion. By the arguments in Rothenberg (1984, p. 898), this will follow if $P(|R_n| > \log^c n) = o(n^{-r/2})$. This will in turn hold if

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

c.f. Bhattacharay and Ghosh (1978, p. 447-448). Since $W_n(z)$ is a smooth function of z, this holds by the tail probability condition imposed on Z_n in the theorem. □

Proof. [Proposition 6] For the SMIL estimator, we employ the general result in Kristensen and Shin (2008, Lemma 8). We first note that conditions (C.1)-(C.3) of Kristensen and Shin (2008) hold under our Assumptions 1-4. Thus, the result will follow if

$$\sup_{\theta \in \Theta} \left| \log \hat{f}_{n,S}(Z_n|\theta) - \log f_n(Z_n|\theta) \right| = o_P(\sqrt{n}).$$

To show this, note that as $n \to \infty$, $f_n(z|\theta)$ is arbitrarily close to $\phi_n^*(z|\theta)$ defined in eq. (9). In particular, for n large enough, $\frac{1}{2}\phi_n^*(z|\theta) \le f_n(z|\theta) \le 2\phi_n^*(z|\theta)$, and similar for its derivatives. We now analyze the kernel density estimator. By standard arguments

$$E\left[\hat{f}_{n,S}(z|\theta)\right] = f_n(z|\theta) + h^2 \frac{\partial^2 f_n(z|\theta)}{\partial z^2},$$

where as $n \to \infty$,

$$\left|\partial^2 f_n(z|\theta)/\partial z^2\right| \le 2\left|\partial^2 \phi_n^*(z|\theta)/\partial z^2\right| = O(n).$$

Similarly we find that the variance component is of order $O_P\left(n/\left(Sh^d\right)\right)$. Combining these pointwise results with standard uniform convergence arguments for kernel estimators (see, for example, Kristensen, 2009), we obtain that

$$\sup_{z \in R^d, \theta \in \Theta} \left| \hat{f}_{n,S}(z|\theta) - f_n(z|\theta) \right| = O_P(nh^2) + O_P\left(\sqrt{\frac{n\log(S)}{Sh^d}}\right).$$

Next, by a first order Taylor expansion, together with the bound $\frac{1}{2}f_n^*(z|\theta) \le f_n(z|\theta)$, it holds for any $B > 0$,

$$\sup_{|z| \le B, \theta \in \Theta} \left| \log \hat{f}_{n,S}(z|\theta) - \log f_n(z|\theta) \right| \le \sup_{|z| \le B, \theta \in \Theta} \frac{1}{2f_n(z|\theta) + o_P(1)} \left| \hat{f}_{n,S}(z|\theta) - f_n(z|\theta) \right|$$

$$\le \frac{C}{\sqrt{n}} exp\left[nB^2\right] \sup_{|z| \le B, \theta \in \Theta} \left| \hat{f}_{n,S}(z|\theta) - f_n(z|\theta) \right|,$$

for some constant $C < \infty$. Since $Z_n = Z(\theta_0) + O_P(1/\sqrt{n})$, we can choose the bound $B = B_n = B_0 n^{-1/2}\log(n)$ for some $B_0 > 0$ and obtain in total that

$$\sup_{\theta \in \Theta} \left| \log \hat{f}_{n,S}(Z_n|\theta) - \log f_n(Z_n|\theta) \right| = O_P(nh^2) + O_P\left(n\sqrt{\frac{\log(S)}{Sh^d}}\right).$$

For the simulated BIL estimator, first note that

$$\hat{\theta}_{SBIL} = \hat{\theta}_{BIL} + E_S(Z_n),$$

where $E_S(z|\theta) := \hat{E}[\theta|Z_n(\theta) = z] - E[\theta|Z_n(\theta) = z]$. In the case where kernel smoothing is used, the same arguments as before together with the result of Kristensen (2009) (see also Creel and Kristensen, 2009), yield

$$\|E_S(Z_n|\theta)\| \le \sup_{|z| \le B_n, \theta \in \Theta} \|E_S(z|\theta)\| = O_P(nh^2) + O_P\left(n\sqrt{\frac{\log(S)}{Sh^d}}\right),$$

while with nearest-neighbor estimators (see Collomb and Härdle, 1986),

$$\|E_S\left(Z_n|\theta\right)\| \le \sup_{|z|\le B_n,\theta\in\Theta}\|E_S\left(z|\theta\right)\| = O_P\left(n\left(\frac{k}{S}\right)^{d/2}\right) + O_P\left(n\sqrt{\frac{log\left(S\right)}{k}}\right).$$

$\square$

Proof. [Proposition 7] First consider the kernel smoothed version of SMIL: Let

$$\hat{\phi}^*_{n,S}(Z_n|\theta) = \sum_{s=1}^{S} K_h\left(Z^*_{n,s}\left(\theta\right) - Z_n\right),$$

denote the kernel density estimator based on i.i.d. draws from the statistic $Z^*_n\left(\theta\right) \sim$ i.i.d.$N\left(Z\left(\theta\right),\Omega\left(\theta\right)/n\right)$. Note that $Z^*_n\left(\theta\right)$ has density $\phi^*_n\left(Z_n|\theta\right)$ as defined in eq. (9). By the same arguments as those used in the proof of Proposition 1, it is easily shown that $\hat{\theta}^*_{MIL} := \text{argmax}_{\theta\in\Theta}\log\phi^*_n\left(Z_n|\theta\right)$ satisfies $\sqrt{n}(\hat{\theta}^*_{MIL} - \theta_0) \to^d N\left(0, J^{-1}\left(\theta_0\right)\right)$. Next, applying the same arguments as in the proof of Proposition 6 on $\hat{f}^*_{n,S}(Z|\theta)$, we obtain

$$\sup_{\theta\in\Theta}\left|\log\hat{f}^*_{n,S}\left(Z_n|\theta\right) - \log f^*_n\left(Z_n|\theta\right)\right| = o_P\left(\sqrt{n}\right).$$

Finally, observe that under Assumptions 2 and 7, $E\left[\sup_{\theta\in\Theta}\left|Z^*_{n,s}\left(\theta\right) - Z_{n,s}\left(\theta\right)\right|\right] = O\left(1/\sqrt{n}\right)$ by dominated convergence. Combining this with the Lipschitz property of K,

$$\frac{1}{\sqrt{n}}E\left[\sup_{z\in R^d,\theta\in\Theta}\left|\hat{f}^*_{n,S}(z|\theta) - \hat{f}_{n,S}(z|\theta)\right|\right]$$

$$\le \frac{1}{\sqrt{n}}E\left[\sup_{z\in R^d,\theta\in\Theta}\left|K_h\left(Z^*_{n,s}\left(\theta\right) - z\right) - K_h\left(Z_{n,s}\left(\theta\right) - z\right)\right|\right]$$

$$\le \frac{D}{\sqrt{n}h^2}E\left[\sup_{\theta\in\Theta}\left|Z^*_{n,s}\left(\theta\right) - Z_{n,s}\left(\theta\right)\right|\right]$$

$$= O\left(1/\left(nh^2\right)\right).$$

This implies that $\sqrt{n}(\hat{\theta}_{SMIL} - \hat{\theta}^*_{MIL}) = o_P\left(1\right)$ under the conditions imposed on S and h. In total,

$$\sqrt{n}\left(\hat{\theta}_{SMIL} - \theta_0\right) = \sqrt{n}\left(\hat{\theta}^*_{MIL} - \theta_0\right) + \sqrt{n}(\hat{\theta}_{SMIL} - \hat{\theta}^*_{MIL}) \to^d N\left(0, J^{-1}\left(\theta_0\right)\right).$$

By similar arguments, the same result can be shown for $\hat{\theta}_{SBIL}$. $\square$

APPENDIX B: AUXILIARY LEMMA

Consider a q-dimensional estimator $\hat{\theta}$ characterized as a root of a random function $m_n\left(\theta\right) \in \mathbb{R}^q$, $m(\hat{\theta}) = 0$. The following Lemma established a higher-order expansion of the estimator:

Lemma 2. *Suppose that $\sqrt{n}\left(\hat{\theta} - \theta_0\right) = O_P\left(1\right)$, and $m_n\left(\theta\right)$ is three times differentiable with its derivatives satisfying:*

$$\frac{\partial m_n\left(\theta_0\right)}{\partial\theta} = D\bar{m}_n + O_P\left(1/\sqrt{n}\right), \quad \frac{\partial^2 m_n\left(\theta_0\right)}{\partial\theta\partial\theta_i} = D^2\bar{m}_{n,i} + O_P\left(1/\sqrt{n}\right),$$

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

for $i, j = 1, ..., q$, where the matrix $D\bar{m}_n \in \mathbb{R}^{q \times q}$ is non-singular. Moreover, for some sequence $C_n = O_P(1)$,

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

in a neighborhood of θ_0. Then the estimator satisfies the expansion in Equation (14) with $\psi_n = D\bar{m}_n^{-1} m_n(\theta_0)$, $A_n = \partial m_n(\theta_0) / (\partial \theta) - D\bar{m}_n$, $B_{n,i} = \partial^2 m_n(\theta_0) / (\partial \theta \partial \theta_i) - D^2\bar{m}_{n,i}$, and

$$Q_1(\psi_n, A_n) = -D\bar{m}_n^{-1} \left\{ A_n \psi_n - \frac{1}{2} \sum_{i=1}^q \psi_{n,i} D^2\bar{m}_{n,i} \psi_n \right\},$$

$$\begin{aligned} Q_2(\psi_n, A_n, B_n) &= -D\bar{m}_n^{-1} A_n Q_1(\psi_n, A_n) - \frac{1}{6} D\bar{m}_n^{-1} \sum_{i,j=1}^q \psi_{n,i} \psi_{n,j} D^3\bar{m}_{n,ij} \psi_n \\ &\quad -D\bar{m}_n^{-1} \frac{1}{2} \sum_{i=1}^q \left\{ \psi_{n,i} D^2\bar{m}_{n,i} Q_1(\psi_n, A_n) + Q_{1,i}(\psi_n, A_n) D^2\bar{m}_{n,i} \psi_n + \psi_{n,i} B_{n,i} \psi_n \right\}. \end{aligned}$$

Proof. We proceed as in Rilstone, Srivasta and Ullah (1996) and Newey and Smith (2004): First, by a third order Taylor expansion,

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

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

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

we obtain

$$\begin{aligned} 0 &= m_n(\theta_0) + D\bar{m}_n(\hat{\theta} - \theta_0) + A_n(\hat{\theta} - \theta_0) \\ &\quad + \frac{1}{2} \sum_{i=1}^q (\hat{\theta}_i - \theta_{0,i}) D^2\bar{m}_{n,i}(\hat{\theta} - \theta_0) + \frac{1}{2} \sum_{i=1}^q (\hat{\theta}_i - \theta_{0,i}) B_{n,i}(\hat{\theta} - \theta_0) \\ &\quad + \frac{1}{6} \sum_{i,j=1}^q (\hat{\theta}_i - \theta_{0,i})(\hat{\theta}_j - \theta_{0,j}) D^3\bar{m}_{n,ij}(\hat{\theta} - \theta_0) + O_P(1/n^2). \end{aligned}$$

Using that $A_n = O_P(1/\sqrt{n})$ and $B_n = O_P(1/\sqrt{n})$, the result now follows by the same arguments as in Newey and Smith (2004, Proof of Lemma A4). □

TABLES

TABLE 1. Dynamic panel data model. Bias. Source for ML and II is Gouriéroux, Phillips and Yu, 2010, Table 2. SBIL, SMIL and II are exactly identified, using the ML auxiliary statistic. SBIL(OI) and SMIL(OI) are overidentified, using both naive and ML auxiliary statistics.

T	N	ϕ	ML	II	SBIL	SMIL	SBIL(OI)	SMIL(OI)
5	100	0.0	-0.199	0.001	0.004	0.002	-0.000	-0.001
5	100	0.3	-0.274	-0.001	0.003	-0.004	-0.001	-0.002
5	100	0.6	-0.362	0.000	0.004	0.002	-0.001	-0.000
5	100	0.9	-0.464	0.000	-0.022	-0.003	-0.000	0.015
5	200	0.0	-0.200	0.000	0.001	-0.003	0.000	-0.002
5	200	0.3	-0.275	-0.010	0.001	0.000	-0.001	0.003
5	200	0.6	-0.363	-0.000	0.001	0.001	-0.001	-0.000
5	200	0.9	-0.465	-0.003	-0.010	-0.007	0.001	0.004
10	100	0.0	-0.100	0.001	0.000	-0.001	-0.001	-0.001
10	100	0.3	-0.135	0.000	0.001	0.003	-0.001	0.001
10	100	0.6	-0.179	0.000	-0.001	0.003	-0.001	-0.001
10	100	0.9	-0.245	-0.001	-0.003	-0.001	0.001	0.006
10	200	0.0	-0.100	0.000	0.000	-0.002	-0.000	-0.002
10	200	0.3	-0.135	-0.000	0.000	-0.002	-0.000	-0.001
10	200	0.6	-0.179	0.000	0.001	-0.001	-0.000	0.001
10	200	0.9	-0.244	0.002	-0.000	0.002	0.001	0.003
20	100	0.0	-0.050	0.001	0.000	-0.002	-0.000	-0.001
20	100	0.3	-0.066	0.001	-0.001	-0.003	-0.000	-0.002
20	100	0.6	-0.086	0.000	-0.000	-0.002	-0.001	-0.001
20	100	0.9	-0.120	0.000	0.000	0.001	0.000	0.002
20	200	0.0	-0.050	0.000	-0.000	0.001	-0.000	0.001
20	200	0.3	-0.061	0.000	0.000	0.001	-0.000	0.001
20	200	0.6	-0.086	0.000	-0.000	0.002	-0.001	-0.000
20	200	0.9	-0.120	0.000	0.000	0.001	0.000	0.001

TABLE 2. Dynamic panel data model. RMSE. Source for ML and II is Gouriéroux, Phillips and Yu, 2010, Table 2. SBIL, SMIL and II are exactly identified, using the ML auxiliary statistic. SBIL(OI) and SMIL(OI) are overidentified, using both the naive and ML auxiliary statistics.

T	N	ϕ	ML	II	SBIL	SMIL	SBIL(OI)	SMIL(OI)
5	100	0.0	0.204	0.057	0.059	0.057	0.044	0.045
5	100	0.3	0.278	0.081	0.065	0.068	0.041	0.043
5	100	0.6	0.365	0.070	0.071	0.069	0.036	0.037
5	100	0.9	0.467	0.076	0.059	0.073	0.033	0.045
5	200	0.0	0.203	0.041	0.041	0.039	0.031	0.030
5	200	0.3	0.277	0.074	0.046	0.048	0.029	0.031
5	200	0.6	0.365	0.050	0.050	0.054	0.025	0.027
5	200	0.9	0.467	0.054	0.046	0.051	0.027	0.034
10	100	0.0	0.104	0.035	0.036	0.034	0.031	0.030
10	100	0.3	0.139	0.036	0.037	0.040	0.030	0.032
10	100	0.6	0.182	0.037	0.038	0.040	0.024	0.027
10	100	0.9	0.246	0.040	0.036	0.046	0.023	0.032
10	200	0.0	0.103	0.025	0.025	0.028	0.022	0.024
10	200	0.3	0.137	0.026	0.026	0.027	0.021	0.022
10	200	0.6	0.180	0.026	0.027	0.029	0.017	0.019
10	200	0.9	0.245	0.028	0.028	0.030	0.016	0.021
20	100	0.0	0.054	0.024	0.024	0.026	0.022	0.024
20	100	0.3	0.070	0.024	0.024	0.023	0.021	0.021
20	100	0.6	0.088	0.022	0.022	0.025	0.018	0.019
20	100	0.9	0.122	0.021	0.020	0.021	0.013	0.015
20	200	0.0	0.052	0.017	0.017	0.018	0.016	0.017
20	200	0.3	0.068	0.016	0.017	0.015	0.015	0.015
20	200	0.6	0.087	0.015	0.016	0.016	0.013	0.012
20	200	0.9	0.120	0.014	0.015	0.016	0.009	0.013

TABLE 3. MA(1) model.

		Bias				RMSE			
T	ψ	ML	CU-II	SBIL(AR)	SBIL(ML)	ML	CU-II	SBIL(AR)	SBIL(ML)
50	-0.95	0.072	0.142	0.109	0.111	0.114	0.190	0.129	0.129
50	-0.9	0.042	0.118	0.077	0.075	0.101	0.172	0.107	0.104
50	-0.5	-0.006	0.058	0.052	0.010	0.141	0.188	0.170	0.142
50	0	0.004	0.001	0.000	0.003	0.157	0.205	0.171	0.151
50	0.5	0.005	-0.044	-0.058	-0.012	0.141	0.184	0.173	0.142
50	0.9	-0.042	-0.066	-0.081	-0.075	0.101	0.142	0.112	0.103
50	0.95	-0.069	-0.088	-0.113	-0.108	0.110	0.145	0.131	0.126
100	-0.95	0.041	0.074	0.053	0.052	0.073	0.104	0.067	0.071
100	-0.9	0.021	0.054	0.025	0.025	0.065	0.091	0.056	0.059
100	-0.5	-0.001	0.024	0.034	0.010	0.092	0.112	0.113	0.093
100	0	0.002	0.001	0.000	0.001	0.105	0.127	0.110	0.103
100	0.5	0.004	-0.019	-0.034	-0.006	0.091	0.109	0.112	0.092
100	0.9	-0.021	-0.021	-0.029	-0.025	0.063	0.087	0.059	0.057
100	0.95	-0.041	-0.046	-0.063	-0.053	0.072	0.093	0.078	0.071
200	-0.95	0.025	0.049	0.030	0.022	0.047	0.071	0.043	0.042
200	-0.9	0.012	0.029	0.003	0.004	0.041	0.059	0.041	0.039
200	-0.5	-0.001	0.009	0.021	0.004	0.063	0.071	0.075	0.063
200	0	-0.002	0.000	-0.000	-0.001	0.073	0.081	0.075	0.072
200	0.5	0.001	-0.008	-0.020	-0.005	0.064	0.071	0.076	0.064
200	0.9	-0.012	-0.007	-0.007	-0.005	0.042	0.067	0.042	0.041
200	0.95	-0.024	-0.024	-0.038	-0.021	0.045	0.068	0.050	0.040

TABLE 4. Nonlinear panel data model.

	$N = 30, T = 5$			$N = 100, T = 5$		
Estimator	Bias	RMSE	MAE	Bias	RMSE	MAE
CU-II	0.056	0.294	0.230	0.024	0.155	0.123
SBIL	0.035	0.260	0.207	0.008	0.142	0.113

TABLE 5. Auction model. Source for II(PML) and II (OLS) is Li (2010), Tables 1 and 2.

		$\theta_0 = 1$		$\theta_1 = 0.5$	
Estimator	Aux. stat.	Bias	RMSE	Bias	RMSE
II	PML	-0.031	0.0534	0.004	0.0230
II	OLS	-0.048	0.0667	0.012	0.0286
SBIL	OLS	0.001	0.0311	-0.001	0.0226
SBIL	extended OLS	0.001	0.0295	-0.001	0.0210

TABLE 6. Fully observed DSGE model with monopolistic competition

Parameter	Lower Bound	Upper bound	True values	Bias		RMSE	
				Prior mean	SBIL	Prior mean	SBIL
α	0.15	0.4	0.33	-0.055	-0.002	0.091	0.006
β	0.95	0.999	0.99	-0.016	-0.000	0.021	0.001
δ	0.005	0.06	0.023	0.010	0.001	0.019	0.001
ψ	1	3	1.75	0.250	0.004	0.629	0.014
ρ	0.85	0.99	0.95	-0.030	-0.017	0.050	0.024
σ	0.005	0.04	0.01	0.012	0.000	0.016	0.001
ϵ	9	13	10	1.000	0.002	1.529	0.033

TABLE 7. Partially observed DSGE model with habit formation, first design

Parameter	Lower Bound	Upper bound	True values	Bias		RMSE	
				Prior mean	SBIL	Prior mean	SBIL
α	0.25	0.4	0.36	-0.035	-0.003	0.056	0.008
β	0.93	0.99	0.95	0.010	0.001	0.020	0.004
δ	0.02	0.04	0.025	0.005	0.000	0.008	0.001
η	0	0.5	0.2	0.050	-0.024	0.153	0.044
γ	1	4	2	0.500	0.220	1.000	0.283
ρ	0.8	0.99	0.85	0.045	-0.005	0.071	0.017
σ	0.01	0.08	0.04	0.005	0.001	0.021	0.004
ψ	NA	NA	3.197	9.854	0.356	22.529	0.530

TABLE 8. Partially observed DSGE model with habit formation, second design

Parameter	Lower Bound	Upper bound	True values	Bias		RMSE	
				Prior mean	SBIL	Prior mean	SBIL
α	0.25	0.4	0.36	-0.035	0.001	0.056	0.006
β	0.93	0.99	0.95	0.010	-0.003	0.020	0.004
δ	0.02	0.04	0.025	0.005	0.001	0.008	0.002
η	0	0.5	0.4	-0.150	-0.033	0.208	0.056
γ	1	4	3	-0.500	0.018	1.000	0.185
ρ	0.8	0.99	0.85	0.045	-0.003	0.071	0.016
σ	0.01	0.08	0.04	0.005	0.000	0.021	0.003
ψ	NA	NA	13.562	-0.511	0.792	22.266	3.052

FIGURE 1. Fully observed DSGE model. Pseudo-priors, true parameter values, and density of SBIL

(A) α

(B) β

(C) δ

(D) ψ

(E) ρ

(F) σ

(G) ϵ
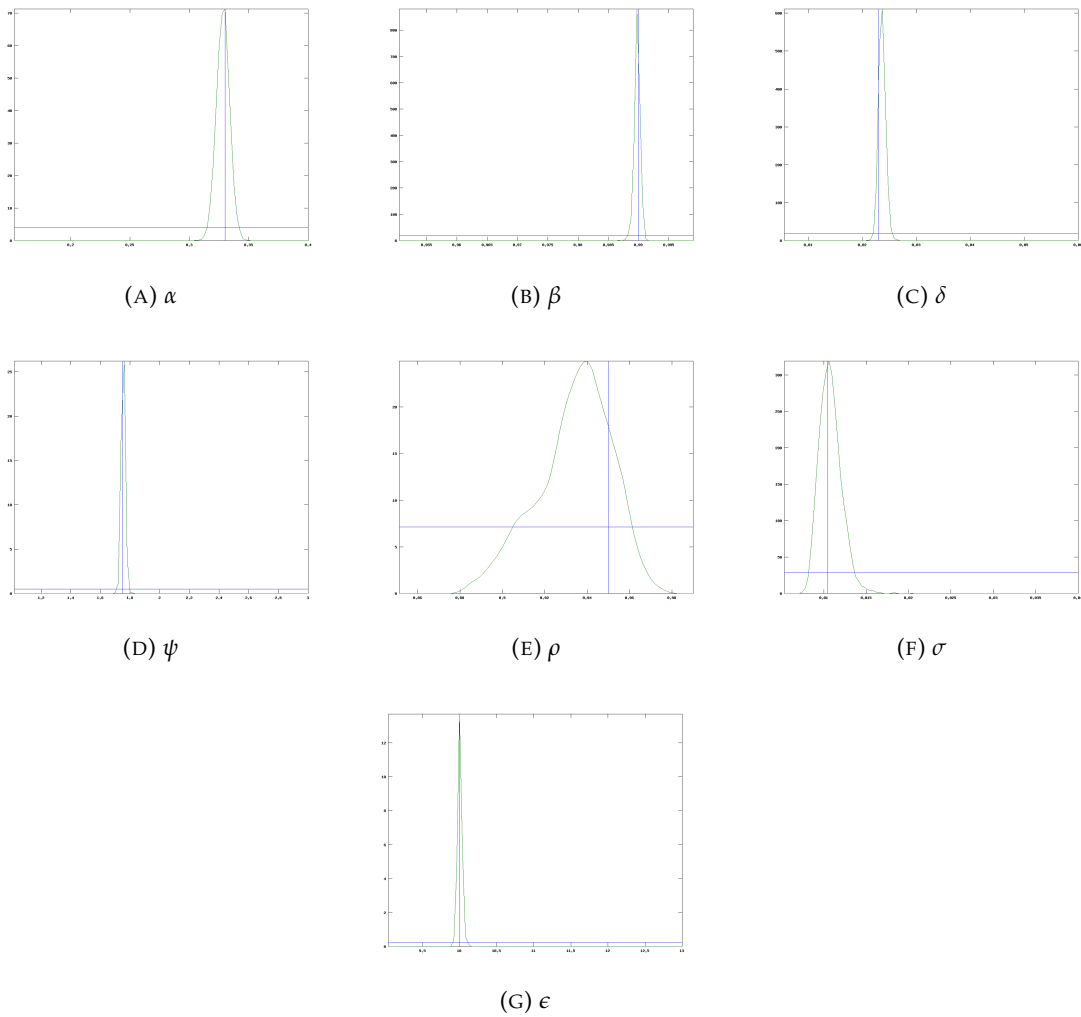
FIGURES

UNIVERSITAT AUTÒNOMA DE BARCELONA AND MOVE

COLUMBIA UNIVERSITY AND CREATES (CENTER FOR RESEARCH IN ECONOMETRIC ANALYSIS OF TIME SERIES, UNIVERSITY OF AARHUS).